30726

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* Janet Elaine King

TITLE OF THESIS/*TITRE DE LA THÈSE* Correlating Verbs With Actions in a Paradigm for Computer Language Acquisition

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/*GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* M Sc

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE DEGRÉ* 1976

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dr. J.R. Sampson

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* Sept. 17, 1976 SIGNED/*SIGNÉ* Janet Elaine King

PERMANENT ADDRESS/*RÉSIDENCE FIXE* 14804 75 Street
Edmonton, Alberta T5C 0Z9

NL-91 (11-73)

INFORMATION TO USERS

THIS DISSERTATION HAS BEEN
MICROFILMED EXACTLY AS RECEIVED

This copy was produced from a microfiche copy of the original document. The quality of the copy is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Canadian Theses Division
Cataloguing Branch
National Library of Canada
Ottawa, Canada K1A ON4

AVIS AUX USAGERS

LA THESE A ETE MICROFILMEE
TELLE QUE NOUS L'AVONS RECUE

Cette copie a été faite à partir d'une microfiche du document original. La qualité de la copie dépend grandement de la qualité de la thèse soumise pour le microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

NOTA BENE: La qualité d'impression de certaines pages peut laisser à désirer. Microfilmee telle que nous l'avons reçue.

Division des thèses canadiennes
Direction du catalogage
Bibliothèque nationale du Canada
Ottawa, Canada K1A ON4

THE UNIVERSITY OF ALBERTA

CORRELATING VERBS WITH ACTIONS IN A PARADIGM FOR COMPUTER LANGUAGE ACQUISITION

by

© JANET ELAINE KING

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE IN COMPUTING SCIENCE

DEPARTMENT OF COMPUTING SCIENCE

EDMONTON, ALBERTA

FALL, 1976

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Correlating Verbs With Actions in a Paradigm for Computer Language Acquisition submitted by Janet Elaine King in partial fulfilment of the requirements for the degree of Master of Science in Computing Science.

.....
Supervisor

.....

.....

.....
External Examiner

Date 12/8/76

To Andy, with love.

ABSTRACT

The undue emphasis by linguists and A.I. Researchers on language structure and generation, at the expense of methods of acquisition, prompted Ian McMaster, in 1975, to propose a Comprehensive Language Acquisition Program (CLAP), emphasizing comprehension. This program differs from its predecessors in its completeness and its realistic combination of modelling and pragmatism. In support of CLAP's feasibility, a Vocabulary Acquisition System was programmed, showing promise for learning input words in relation to their corresponding referents in a static environment.

A Verbal Acquisition Module is now proposed as an extension of VAS to include events in its environment, with representations patterned after Schank's conceptualizations. The module is seen as another step towards the fulfillment of the larger language acquisition system, as it was proposed, or with modification.

After a brief introductory chapter, previous related research in psycho-linguistics and computer science is presented. CLAP is discussed as a first step towards the specification of a complete acquisition model, containing all the required ingredients though lacking sufficient detail for programming at this time.

Extending processes similar to those employed by VAS,

VAM correlates words with concepts relating to actions and events, whose representation is inspired by Schank's proposed structures for actions and events.

Some possible extensions of this work and further research prospects are mentioned in the final chapter.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to Dr. J.R. Sampson, whose encouragement and time, as well as his patience in anticipating the results of this course of study, were above and beyond the call of duty. And for his inspirational ideas and suggestions, I would like to express my gratitude to Ian McMaster.

I would also like to thank the National Research Council of Canada, and the Department of Computing Science at the University of Alberta, whose financial assistance enabled me to undertake this work.

Finally, my experimental data I owe to Mrs. Karen Jerat, who showed great imagination in pretending the BLOCKs world was interesting enough to discuss with a child.

TABLE OF CONTENTS

1 INTRODUCTION 1
1.1 A Comprehensive Language Acquisition Program 3
1.2 The Next Step 4
1.3 Overview 6

2 PSYCHOLINGUISTICS 8
2.1 The Genesis Of Acquisition 10
2.1.1 Lad 11
2.1.2 The Linguistic Environment 12
2.1.3 How The First Words Are Acquired 14
2.2 A Conceptual Structure For Actions 16
2.2.1 The Acts 18
2.2.2 Conceptual Dependencies 20
2.2.3 Evidence In Support Of Conceptual Structures 22
2.3 Missing Parts 24
2.4 Criteria For A Model 25

3 COMPUTERS AND NATURAL LANGUAGE 28
3.1 Other Models Of Language Acquisition 29
3.1.1 Language Acquisition By Hypothesis Testing 29
3.1.2 A General Problem Solving Model Applied To Natural Language Acquisition 35
3.1.3 Deep Structures As Card Configurations 38
3.1.4 Another Problem-solving Robot 42
3.2 Clap 42
3.2.1 Components Of The System 43
3.2.2 The Learning Strategies 47
3.2.2.1 Strategy 1: Segmentation And Meaning Association 48
3.2.2.2 Strategy 2: Linear Ordering 51
3.2.2.3 Strategy 3: Structural Generalization 52
3.2.2.4 Strategy 4: Conflict Resolution 54
3.2.2.5 Strategy 5: Using Discourse 55
3.2.3 Evaluation 56
3.2.4 A Test Of Plausibility 59

4 A VERBAL ACQUISITION MODULE 64
4.1 Conceptual Structures For Vam 65
4.1.1 Components Of The Actor-action Conceptualization 68
4.1.2 Cases Of The Action 70
4.1.3 Relevant Acts And Their Inferences 72
4.2 Objects In The Focus 77
4.3 Building A Focus 78
4.4 Evaluating The Associations 81
4.5 Preliminary Experiments And Results 83
4.6 Correlation Results 87

5 CONCLUSIONS 94
5.1 Extensions To Vam 95
5.2 Towards The Implemetation Of Strategy 1 98

REFERENCES.....100

APPENDIX A: THE CORPORA FOR VAM 103

APPENDIX B: PRELIMINARY RESULTS WITH AN ALTERNATIVE CORRELATION FUNCTION 108

APPENDIX C: ABBREVIATIONS USED 113

LIST OF FIGURES

2.1 Physical ACTs 19

2.2 Relations 21

2.3 Case Links 22

2.4 Possible Components of Schank's Actor-Action Conceptualization 23

3.1 Flow Design of Kelley's Language Acquisition Model ..32

3.2 Components of CLAP 44

3.3 Parsing in Strategy 1 50

4.1 The Environment for VAM 67

4.2 Actor-Action Conceptualizations as Seen by VAM 73

4.3 Results Using Corpus 1 89

4.4 Results Using Corpus 2 90

4.5 Results Using Combined Corpus 91

B.1 Correlations for the Word Block 109

B.2 Correlations for the Word B8 109

B.3 Correlations for the Word Moved 110

B.4 Correlations for the Word VAM 110

B.5 Correlations for the Word Triangle 111

INTRODUCTION

Artificial Intelligence (A.I.) is the field within computer science concerned with the mechanical performance of tasks previously assumed to require human intelligence.

> Problem-solving, perceptual, and linguistic competence all seem to depend upon prior knowledge of the task domain being present a priori in the computer. Hence the question of how a computer can acquire, represent, and make use of the knowledge of the world, is the main problem of AI. <Raphael, 1973, p. 9>

Computers have been programmed to answer questions, play games, and solve problems when provided with the necessary information and means of using it. Some such models of intelligent activity have been more elegant than others, the more impressive ones appearing to understand and think rather than to trivially manipulate the input to produce the output.

Modelling, in the strict sense of the word, would require a precise definition of the process to be modelled. Human cognition, still far from being completely understood,

cannot be adequately modelled. The notion of "model" must require only the simulation of performance, not processes. The current state of psychological expertise need not discourage work in artificial intelligence. Although intelligence in any non-human species is measured by human standards, imitation of such behavior by machine, in any manner which is sophisticated enough to be termed intelligent, can be not only a useful tool for experimenting with machine irtelligence, but may well provide clues to how people think.

The ability to communicate through language is traditionally thought to be one of the distinguishing traits of man. Computers have been used to test linguistic transformational theories and tc implement would-be mechanical translators. In the area of A.I., Winograd <1971,1972> has developed SHRDLU, a robotic system which demonstrates a remarkable ability to relate verbal input to its knowledge of the universe, represented by a world of various geometric BLOCKS which it can manipulate and "see". The presence of an environment gives substance to meaning. Only when language expresses meaning does it interest the A.I. Researcher; and only in the context of an environment can there be subject material for meaningful expression.

Winograd's system begins with a pre-defined language. Except for a limited number of recent studies <see Chapter

3>, little A.I. research has been devoted to the behavioral processes involved in acquiring the use of language as a tool for communication. The process of language learning, of how and why a child can learn to understand and speak, is one of the most fascinating of all aspects of language use. A complete model of language performance will necessarily include a theory of the process of language learning, as has been emphasized by Schwarcz <1967>.

1.1 A Comprehensive Language Acquisition Program

A Comprehensive Language Acquisition Program (CLAP) has been proposed by Ian McMaster <1975> as a system for acquiring natural language, given a subset of descriptive sentences input from a terminal, an internal representation of an environment, and a means of receiving inputs consisting of actions on the environment (a), approval or disapproval (r), an utterance (u), a stimulus to output (s), or a variety of combinations of these <McMaster, 1975, pp. 89-90>. Eventually, the system learns to converse.

The acquisition system is based on five sequential learning strategies. The emphasis is on learning to understand, production being dependent on comprehension. The first strategy involves learning to segment lexical strings and to associate meaning with these segments. Once

CLAP has reached the point where it understands more than one word of a given sentence, it can attach some import to word order (strategy 2). Structural Generalization, Conflict Resolution, and the Use of Discourse complete the five strategies, which will be described in Chapter 3.

1.2 The Next Step

CLAP is a proposal for a program which would acquire language. To demonstrate the feasibility of programming such a model, McMaster <1975> wrote VAS (Vocabulary Acquisition System). VAS associated a word and its corresponding concept (which was an entity, attribute, or relation) through weighted association links. However, VAS could not manipulate objects, nor could the user. If object movement is allowed within a scene, perhaps a system patterned after VAS could also acquire action-concept associations.

A major problem in our design of a verbal acquisition module (VAM, to conform with the trend to give pet names to such projects) is the internal representation of an action on the environment, and the event surrounding its occurrence. Roger Schank <1973b> suggests that all conceptualizations, or sentences, can be reduced to a structure built around primitive actions. This claim will

be explored further in Chapters 2 and 4.

We shall assume that such actions as the child can recognize nonlinguistically provide a conceptual structure onto which a lexicon can be mapped. Those environmental modifications which were too subtle for the child to notice will be ignored until he has enough experience to understand the concept. Before this, however, the structure exists, according to Schank, in which slots will later be filled.[1]

Changes in the environment, as well as concepts of objects and their attributes, will all be part of VAM's focus of attention, which contains those portions of the environment noticed by VAM at a particular time. This list of data base elements will be built by the programs required to carry out a particular event. Some programs will correspond to the conceptualization level of Schank's schema. Others will represent the primitive actions chosen for VAM's environment. Objects involved will fill slots in the list, while the relationships will be derived from the relational position of objects to each other. VAM will link the input lexemes of the utterance to these concepts.

---

[1] "...the conceptual apparatus that underlies adult language is present in a child before he has finished his first year of life. It is this conceptual apparatus that guides language learning and in fact facilitates the infant's handling of the world in general." <Schank, 1973c, p. 1>.

Although an attempt will be made to make this system psychologically and linguistically plausible, it should not be construed to be an exact model of human cognitive processes, for reasons stated above.

1.3 Overview

Psychology, linguistics, and, to some extent, computer science have made contributions in the area of child language acquisition. Chapter 2 is concerned with relevant discoveries in these disciplines. After looking at a few of the more interesting language acquisition models, Chapter 3 will examine CLAP's components and strategies. In Chapter 4, the Verbal Acquisition Module will be described as a continued partial implementation of the first of CLAP's strategies. According to CLAP's proposed scheme of implementation <McMaster, 1975>, the number of times a word has been used in connection with a concept, in conjunction with the number of times the word and concept have been used independently, will be considered in determining whether and to what extent a word should have a concept as its meaning.

The major contribution of this work will be the critique of prior work and its generalization to the acquisition of verbs. Input will include utterance-action

pairs, which were absent from the input to VAS. Finally, the results of programming this component will be discussed, with consideration of problems, suggestions for improvement, and directions for further research.

2

PSYCHOLINGUISTICS

The linguist's approach to acquisition has been predominantly concerned with the acquisition of syntax, as evidenced in production samples. Little material is available on holophrasis, the one-word utterance. It is usually discarded as uninterestingly trivial in structure. Yet the child's first word probably holds all the intended meaning of an adult sentence <Brown, 1973>.

Language acquisition research tends to ignore the presence of meaning; at best, its existence is acknowledged in a passing comment, to the effect that there is not enough known about the role of meaning in language learning. As systematic as language may appear, it has no significance without its associated meaning. Although syntax is important, it is a late component in a long process which begins at (or possibly even preceding) birth <Schank, 1973a>.

It is our belief that meaning is not only a valuable tool, but a necessary component in the acquisition process. Instead of being preoccupied with the acquisition of language structure, we agree with Macnamara <1972> that the first meanings are acquired independent of language, and that the first words are then related to the meanings of the utterance through the situational context. A parent's intentional meaning - through expression, tone of voice, and gesture - can often be understood before the associated utterance. The meanings can refer to the physical environment (the only reference in our model), the feelings of the child, his ideas or concepts, and his attitudes regarding truths <Macnamara, 1972>.

Furthermore, understanding of language precedes production. One need only observe a non-speaking infant to realize that he can understand scoldings and carry out commands. As pointed out by Reeker <1974>, what is important is the child's mental grammar, which is the only true test of competence. Unfortunately, the inaccessability of this grammar through performance data, coupled with its rapid advancement and the variations between the grammars of children at a given age, require the linguist to rely on the data he has. The first words spoken provide one of the easiest clues for determining which words are associated with correct meanings, with a minimal amount of

interpretational bias. Linguists also use this data to infer a grammar of the child's language as it stands.

The experimenter can only infer linguistic competence from production performance. There are problems with this approach. It is too easy to become subjective, imposing an adult interpretation of the child's language. It is also difficult to elicit required data to prove or disprove any hypotheses about language acquisition. If a word or concept is not produced, is it because it is not yet mastered, because it was not detected at the right time, or because there was simply lack of a need to use it? Yet we can make some conjectures of what a child has internalized by observing his actions and listening to his speech.

2.1 The Genesis of Acquisition

Among questions pertinent to verbal acquisition which have been raised by both linguists and psycholinguists are three whose answers are still disputable: (1) what is the nature of what we will henceforth loosely refer to as the language acquisition device (LAD)[1] which the child brings to the language-learning situation; (2) what is the

---

[1] This is not to be equated with the innate mechanism described by Chomsky <1965>.

influence of the linguistic environment on the child's learning of a language; and (3) what strategies precede the first spoken sentences? These will each be explored in the following sections.

2.1.1 LAD

Not much of a concrete nature can be said about the language acquisition device.[2] There are two points of view concerning the type of entity the LAD might be. On the basis of putative language universals, some suggest that the child has a built-in sense of the hierarchy of grammatical categories and knowledge of the basic grammatical relations. On the other hand, others prefer to regard only the procedures and inference rules as universal <Bowerman, 1973>. Anderson and Bower <1973> suggest looking to evolution as a clue:

> Both in the evolution of man and in the development of the child, the ability to represent perceptual data in memory emerges long before the ability to represent linguistic information. We believe that language attaches itself to this underlying conceptual system designed for

---

[2] "...the question of the innate apparatus which the child brings to the language-learning situation is subject to a wide spectrum of possible interpretations, none of which seem to be decisively excluded by the range of relevant data which is presently at our disposal." <Derwing, 1970, p. 82>

> perception.... Indeed, it could be argued that natural languages can be learned initially only because their organization corresponds (at least in the simple cases) to the perceptual organization of the referential field. <p. 154>

Whether the LAD be innate, or learned in the prelinguistic psycho-motor stage <Flavell, 1963>, the human child possesses the necessary tools which enable him to learn to use language to communicate with other people. The cognitive structures which may still be undeveloped would limit what the child is capable of learning. If we accept the hypothesis that the LAD, or process itself, is modified with experiencial feedback, as the language itself changes, we have the beginning of a viable theory of learning in general.[3]

2.1.2 The Linguistic Environment

The question arises as to the type of speech which enables a child to learn to understand the language. The child, under normal circumstances, receives a noisy input, full of inconsistencies, incorrect grammar, and make-believe words such as the French "dodo". Often a parent attempts to

---

[3] "...it is maintained that as new structures are obtained, the actual learning mechanism is altered...." <Reeker, 1974, p. 37>

speed up language learning through the use of "baby talk". Extreme simplification or distortion are unnecessary to the learning process, and possibly even a hindrance. A child has internalized his own language system at a given time. For him, noise will be filtered out while the useful input, slightly beyond his present grammar, will be noticed. Those sentences which can be mostly understood but which add something new will provide the most useful data for progression.[4]

Extensions of a child's speech, in which an adult fills in the rest of the skeleton of meaning indicated by the child, is also of questionable necessity, though few would dispute that it helps in learning how to fully express an idea.[5] (This does not refer to syntactic or grammatical corrections, which have little or no value <Brown, 1973>.)

The attention and care given a child certainly influence his acquisition <Deese, 1970>. Addressing the child directly, especially if an important agent in

---

[4] "...what the child wants are instances of data that will either serve to confirm (or infirm) previously acquired constructs or are examples that will bear on the next step in acquisition, but not evidence for constructs which the child will not be ready to acquire until much later." <Kelley, 1967, p. 83>

[5] "Expansions...may present suitable conditions for children to discover the local expression of linguistic universals and do so in a way that imitation and practice do not." <McNeill, 1966, p. 74>.

satisfying his wants, draws him into the center of attention, encouraging performance. This also draws his attention to what is intended in the way of meaning as well as to what is being said <Richelle, 1971>.

Approval is a form of expansion and a motivation to reinforce a well-received utterance. Since the goal of language is to communicate, the child receives a form of approval when he is understood. As the child meets a larger number of people, the conditions for being understood become more stringent and the child's idiolect must conform more and more to the language of a larger community <Kelley, 1967>.

2.1.3 How the First Words are Acquired

The sensori-motor period provides the child with concepts, to which he can later attach spoken input <Nelson, 1974>. The child continues to be drawn to that which is new. This is important when learning the terms associated

with objects.[6] Movement, as well as new objects, elicits a strong orienting response.[7] A child's attention is drawn not to the objects involved in the action, but to the action itself <Brown, 1973>.

Carroll <1964> lists three sequences of development: cognitive; the capacity to discriminate and comprehend speech; and the development of the ability to produce speech sounds that conform more and more to adult speech. We are interested primarily in the second sequence, but all three are closely interrelated. As mentioned above, the cognitive development determines which concepts can have referents, and therefore meanings, closely associated with them.[8] The tuning of an individual's language to that of the community is more important in learning syntax than in learning the

---

[6] "...'even at the age of twelve to thirteen months [word-image] connections can be formed under some conditions after a single reinforcement. The most important condition for developing these connections is the presence of an intense orienting reaction to the named image, as in the case when a new image is placed among other images whose names the child already knows.'...." <Slobin, 1966, p. 145>

[7] "The one outstanding general characteristic of the early words is their reference to objects and events that are perceived in dynamic relationships: that is, actions, sounds, transformations - in short, variation of all kinds." <Nelson, 1974, p. 269>

[8] "...the nature of sensory-motor intelligence severely constrains the range of relational meanings expressed, including even the child's notions of possessive relations between persons and objects, of attributes of objects and his use of apparently 'experiential' verbs." <Edwards, 1974, p. 2>

first few words.

The child's first words are usually overgeneralized. Macnamara <1972> suggests that this displays the child's tendency to take short-cuts. His strategy is to associate a word naming an object with the entire object, rather than with its parts or attributes. Thus, names for entities are often learned first. Often, as in the case of "Dada" <Brown, 1973>, names refer not only to the entity, but to everything associated with it.

Once a child grasps the names of several objects exemplifying an action or a changing state, these concepts can be associated with the appropriate word. These concepts, too, are over-generalized. Slobin <1966> cites the inability to separate the spatial relation "under" from the act of placing one object beneath another.

Permanent attributes are learned last. A good reason would be the difficulty in indicating to a child something such as color, since it is taken for granted and does not elicit an orienting response. Furthermore, how can a child differentiate, without a large number of examples (and, possibly, verbal explanation as well), among such attributes as color, shape, size, etc?

2.2 A Conceptual Structure for Actions

Roger Schank is concerned with presenting a language processing model which accurately reflects human language understanding. The theory is not completely implemented, though parts of it have been programmed by Schank and his students. Schank presents a useful structure for storing an event. It is this structure which will become important to the design of a conceptual structure for VAM. The theory centers around an actor-action conceptual base and a network of concepts (not words) upon which human cognitive processes act.

The conceptual processor consists of the following components: a dictionary of possible realizates of a word; a dictionary of ACTs used in selecting a correct meaning structure within a semantic environment; conceptual expectations; a list of conceptual dependencies; and heuristics for finding the main nominal and action.

Two levels of representation are syntactically and semantically based, repectively. The sentential level deals with utterances encoded within a syntactic language structure. The conceptual level consists of conceptualizations, which are concepts plus the relations among them. It is important not to confuse conceptual with sentential structures. The conceptual structure is mapped

onto the sentential structure as a one-to-many relation. In other words, there may be more than one sentence which expresses the same conceptualization. There is no direct relationship between the ACTs and verbs, either. Not every verb of the sentential structure will be expressible as an ACT, but may describe or modify instead. An example is the verb, hurt, which really describes a resultant state. Also, not every conceptual structure will be expressed verbally, but may be understood, as in the case of the subject of an English imperative.

Schank divides concepts into three types: nominals, entities which can be visualized, or picture producers (PP's); actions, which an animate object must be capable of applying to an object (ACT's); and modifiers, which only have meaning as Picture Aiders (PA'S) or Action Aiders (AA's). (Schank has more recently added two additional concept types, time and location <Schank, 1973d>.)

2.2.1 The ACTs

As of 1973, Schank had reduced the expression of all actions to a set of fourteen primitive actions, each having associated actions or states which can be inferred from its occurrence <Schank, 1973b>. He places these into four categories: instrumental, physical, mental and global.

2.2.1 The ACTs

Instrumental ACTs include SMELL, SPEAK, LOOK-AT, and LISTEN-TO, all of whose objects (smells, sounds, physical objects, and sounds) are used as instruments of the verb. Physical ACTs require a physical object, some more stringent than others (see figure 2.1).

ACT	MEANING	OBJ REQUIREMENTS
PROPEL	to apply a force to	any object
MOVE	move a body part	bodypart
INGEST	take something inside you	any object
EXPEL	take something from inside you and force it out	previously ingested object
GRASP	to grasp	size limit

Figure 2.1: Physical ACTs

Mental ACTs are CONC, meaning to focus attention or perform a mental process on, and whose object is not a concept, but a conceptualization; MTRANS, which describes a change in the mental control of conceptualizations to and from the conscious mind; and MBUILD, to build an internal thought combination.

The global ACTs are PTRANS, involving a change in the physical locations of an object, and ATRANS, indicating a change in ownership.

The verb-ACT dictionary itself is a list of conceptual structures associated with each syntactic and semantic environment a word can have. Each dictionary entry lists each possible structure for a word, indicating the slots to be filled for each word sense.

2.2.2 Conceptual Dependencies

The conceptual structures indicate the conceptual dependencies and can be illustrated through the use of links. The dependencies are recognizable as the condition where the dependent item predicts the existence of the governing one. There are basically two types of links, relation (figure 2.2) and case (figure 2.3), corresponding to two levels of hierarchy:

> ...what makes a relation different from a case....is that a case is part of an underlying ACT and is predicted by that ACT. A relation is a rule for connecting different conceptualizations. Thus, relations serve as connectors within a memory whereas other types of dependency connect things within a conceptualization. <Schank, 1973b, p. 205>.

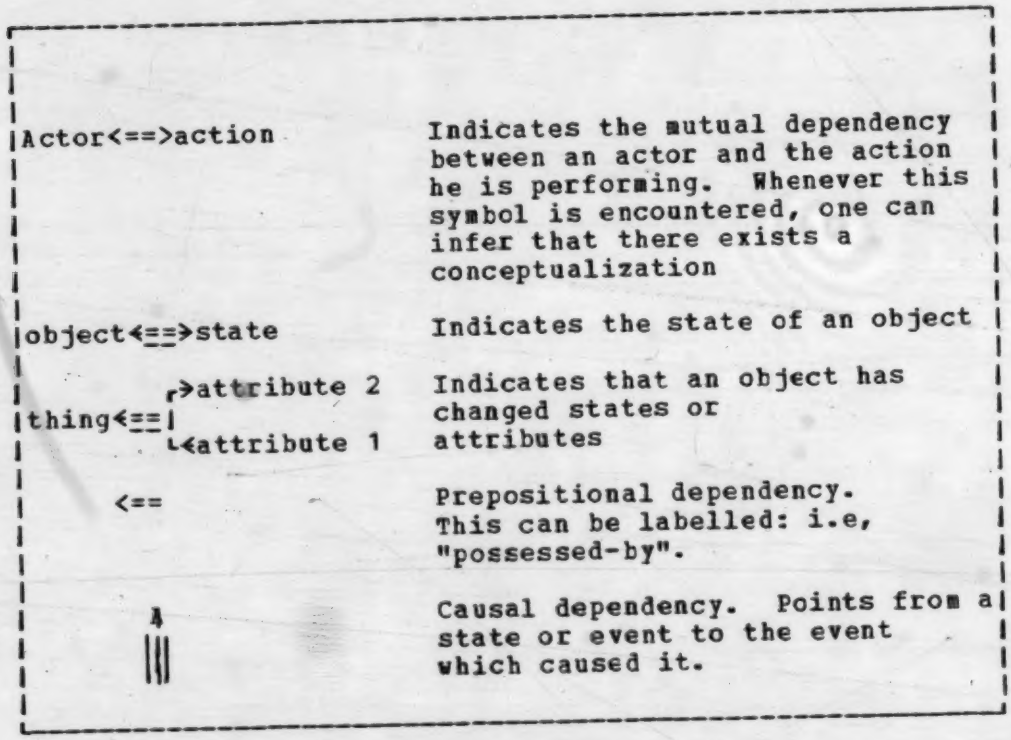

Actor<==>action
Indicates the mutual dependency between an actor and the action he is performing. Whenever this symbol is encountered, one can infer that there exists a conceptualization
object<==>state
Indicates the state of an object
thing<==
>attribute 2
<attribute 1
Indicates that an object has changed states or attributes
<==
Prepositional dependency. This can be labelled: i.e, "possessed-by".
Causal dependency. Points from a state or event to the event which caused it.

Figure 2.2: Relations

Both objects and actions can have attributes. Entire conceptualizations can also have times and locations (see figure 2.4). Limitations are placed on what kinds of things qualify as actors, objects, and cases:

> ...any action that we posit must be an actual action that can be performed on some object by an actor. Nothing else qualifies as an action and thus as a basic ACT primitive. The only actors that are allowed in this schema are animate. <Schank, 1973b, p. 3>

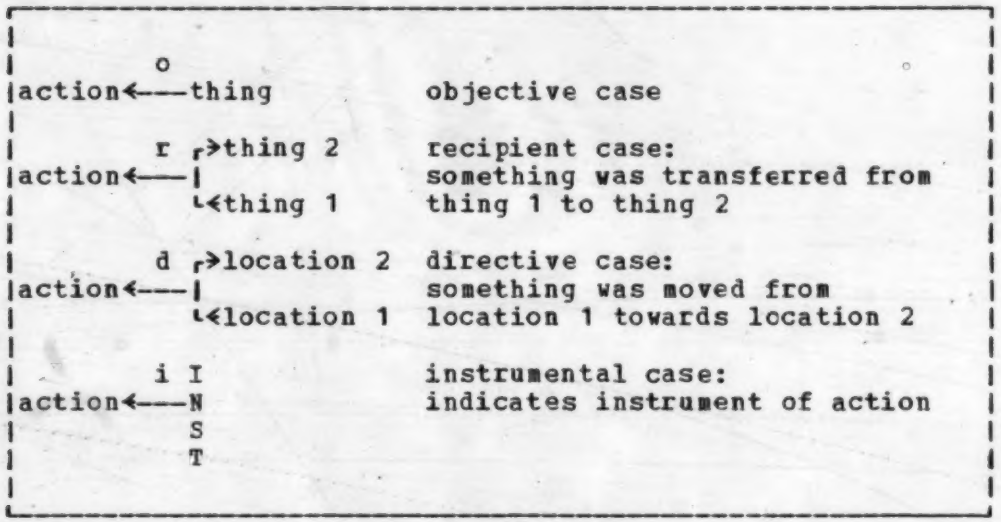


Figure 2.3: Case Links

2.2.3 Evidence in Support of Conceptual Structures

In a 1973 study, Schank <1973a> shows evidence which indicates that, if the Conceptual Dependency theory is assumed to be a reasonable model for conceptual structures underlying natural language, such structures are nearly completely formed before spoken language is present in the child. He asserts, furthermore, that such structures form the basis for language learning when it takes place.

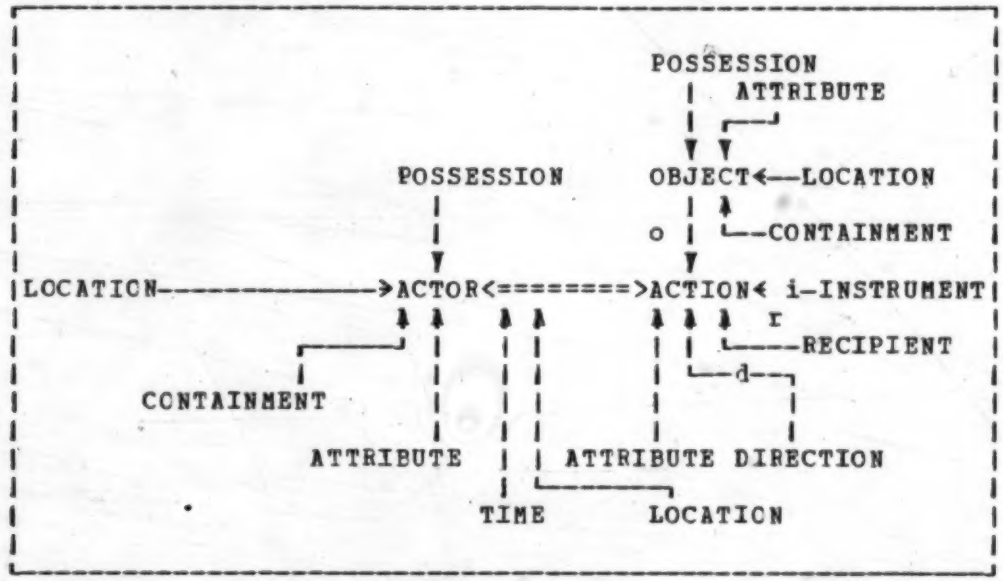


Figure 2.4: Possible Components of Schank's Actor-Action Conceptualization
(My Diagram)

This conclusion was drawn from experiments with two children of ages 0-1 and 2.2-2.4, respectively. For the older child, utterances were used as evidence that certain ACTS, cases and relations were present in the intended meaning. The younger child, Schank's daughter, Hana, was observed closely for demonstration of the intention to perform an action. Care was taken to differentiate between a simple action and the planning which indicates an understanding of the action as a concept. Schank concludes that, by age 1, Hana was aware of all ACTS but EXPEL, SMELL, MBUILD, and CONC.

It is emphasized that this study was not intended to

justify the psychological validity of the Conceptual Dependency theory. It does indicate which conceptual structures a child might have internalized at the time she first begins to use language. Of course, a child begins understanding before speaking, so the beginning of language learning may occur simultaneously with the learning of concepts in the environment.

2.3 Missing Parts

From the information presented in this chapter, it should be clear that more needs to be known in the area of human language acquisition before the process can be modelled. Kelley <1967> outlines areas for research:

1. Techniques need to be developed for interpreting which functional relations are being expressed in child speech.

2. We need to establish how a child's speech develops over time.

3. Finally, we need to establish which developmental sequences, if any, are invariant over all languages.

Further areas should become evident over the next two chapters, some of which will be discussed in depth. In a child's environment, what is the role of attention and salience in the ability to learn to communicate associated concepts? It appears likely that there are levels of

attention which help determine the probability of learning the corresponding word or words. If a concept appears more often in the environment, maybe it is more likely to be noticed than those appearing only once; on the other hand, if it is novel, its importance would be greater to the child than concepts with which he is familiar. Those concepts drawn into the child's focus of attention are those most likely to be associated with co-occurring words.

To determine many of these answers, the emphasis of research needs to be on relating form to meaning, rather than on relating forms to each other <McMaster, 1975>.

2.4 Criteria for a Model

With so many gaps in our understanding of human language acquisition, what can the computer model builder hope to accomplish? A criterion, which seems basic to the definition of "model" is that it perform like a child <Schwarcz, 1967>. More specifically, we feel it should have the following features:

1. Realistic input. i.e., it must not require correct parses or expansions.

2. A realistic environment for reference.

3. Some progression with maturity, whether it be a mere dependence on previous accomplishments, or an imposed maturing process accompanied by the emergence of cognitive capabilities.

With child-like performance as the overall goal, equally suitable lists of alternative objectives could be developed.

Computer modelling, by its nature, imposes restrictions on the nature of the linguistic and conceptual environments and of the Language Acquisition Device. The sophistication of equipment available and the operating system for implementation limit the model to less ambitious projects than brain simulation. For instance, the linguistic input at present cannot be aurally perceived, but must be input orthographically.[9] Furthermore, only one type of input may be received at a time, so that an utterance and an event cannot take place simultaneously (though we can handle them as if they co-occurred). There are no accompanying gestures, pitch modulations, pauses, or emotions other than goals or desires built in to the program.

The non-linguistic environment can be input as artificial "vision", consisting of a camera for a robot, a display screen, or a description of positions in three-dimensional space.

The cognition of the computer does not approach the complexity of the human brain, nor are the processes carried

---

[9] The work of Miller <1974> does indicate a future solution to this problem.

out in a like manner. Experience, for practical purposes, must be artificially stored in the model. The LAD, too, must be built-in, rather than learned. It should include any concept the programmer wishes the system to learn to recognize.

On the basis of Schwarcz's <1967> proposal for an acquisition component in a language model, McMaster <1975> develops a Comprehensive Language Acquisition Program. This system is the framework for the development of a Verbal Acquisition Module.

Schwarcz <1967> proposes that such a model follow an order of stages which consist of 1) recognizing sounds as lexical items, 2) associating such items with referents, 3) linear ordering, 4) generalization into classes, and 5) learning equivalent modes of expression.

In the following chapter, acquisition systems will be examined against the above criteria. Work along the lines of verb processing and acquisition will be incorporated, as needed, in the development of the component, VAM, which learns to associate the first action concepts with their lexical representations.

3

COMPUTERS AND NATURAL LANGUAGE

Natural language, for our purposes, is a dialect of English, although we hope the application could include any language used for human communication. Ambiguities, abstractions, untruths, conjectures, and idioms challenge one to imitate on a machine a process resembling language understanding. Even more difficult is simulating the development of the abilities to use language to interpret speech and to communicate ideas.

Psychological theory has no complete explanation for what happens when a person interprets an utterance, much less how he becomes capable of deriving an appropriate interpretation of the intended meaning from the verbal input. Neither has linguistic theory accounted for the development of the relationship between meaning and structure. Computer science has had only modest success in the area of explicating the acquisition phenomenon. McMaster <1975> is the first to have outlined a truly

comprehensive acquisition model. Other attempts have been made, the more important of which will be reviewed here. Their main problems were concentration on a single aspect of language learning and lack of enough realism to make them impressive simulations of any human process <see McMaster, 1975, chapter 3>.

3.1 Other Models of Language Acquisition

McMaster <1975> and McMaster, et. al., <1976> review prior models in preparation for the presentation of CLAP. Dependency analysis and Jordan's <1972> MEchanical Translator and Question Answerer were both discussed in terms of their contributions and inadequacies, and will not be covered here. On the other hand, Kelley <1967> and Harris <1972> have both made rather significant steps in the right direction and, therefore, deserve special consideration. In addition, the robots of D. Block, et. al. <1975>, and Reeker <1974> deserve mention as new developments in this area.

3.1.1 Language Acquisition by Hypothesis Testing

Kelley <1967> attempts to propose a realistic model of language learning which emphasizes the learning of syntactic

structures. The model is based on the theory that

> 1. Syntactic acquisition is based on the child's comprehension process.
>
> 2. Hypothesis-testing is the mechanism for syntactic acquisition.
>
> 3. Semantics is central to syntactic acquisition.
>
> 4. Only meaningful sentences provide data in acquiring syntactic competence. <Kelley, 1967, pp. 148-149>

As we shall see below, we can accept all but the second point of his theory. Kelley's notion of hypothesis testing is a questionable part of language acquisition. Sentences which are unacceptable as syntactic data (point 4 above) are those which are radically ungrammatical or so complex that the model cannot at least partially interpret the sentence.

Input data in Kelley's system is a set of sample sentences generated by the system itself from a phrase structure grammar. The system has access, through the comparator component, to the correct interpretation <see Figure 3.1>, which is the parse. The system does not label selectional restrictions, so it is unable to learn such semantic features as +/- animate.

The model tests two kinds of hypotheses: initial and generated. The initial hypotheses are programmed to correspond to the three acquisition stages described below.

Hypotheses about lexical categories are used to classify words by semantic definitions of categories. Hypotheses about funtional relationships include a semantic definition of the relationship and knowledge of which other relationships are required in the presence of the functional relationship. These relations are used to interpret the sentence.

In generating hypotheses, the system singles out possibly important properties as candidates. If possible, data is gathered to support an hypothesis. Otherwise, the hypothesis will atrophy out of the system from lack of use.

A developmental time scale determines when successive stages will be initiated. Each such stage generates a different set of initial hypotheses to be tested against sentences <see Figure 3.1>. Once an hypothesis is sufficiently well-confirmed, it is considered to be an acquired grammatical construct.

Stage I hypothesizes only that a single word will refer to a concrete reference and be placed in the "thing" category. In Stage II, "things" and "actions" are separate categories. Functionally, a reference will consist of the "concrete referent" plus "modifier of sentence". The system generates its own hypotheses about the order of funtional relations and which categories can serve in which relation.

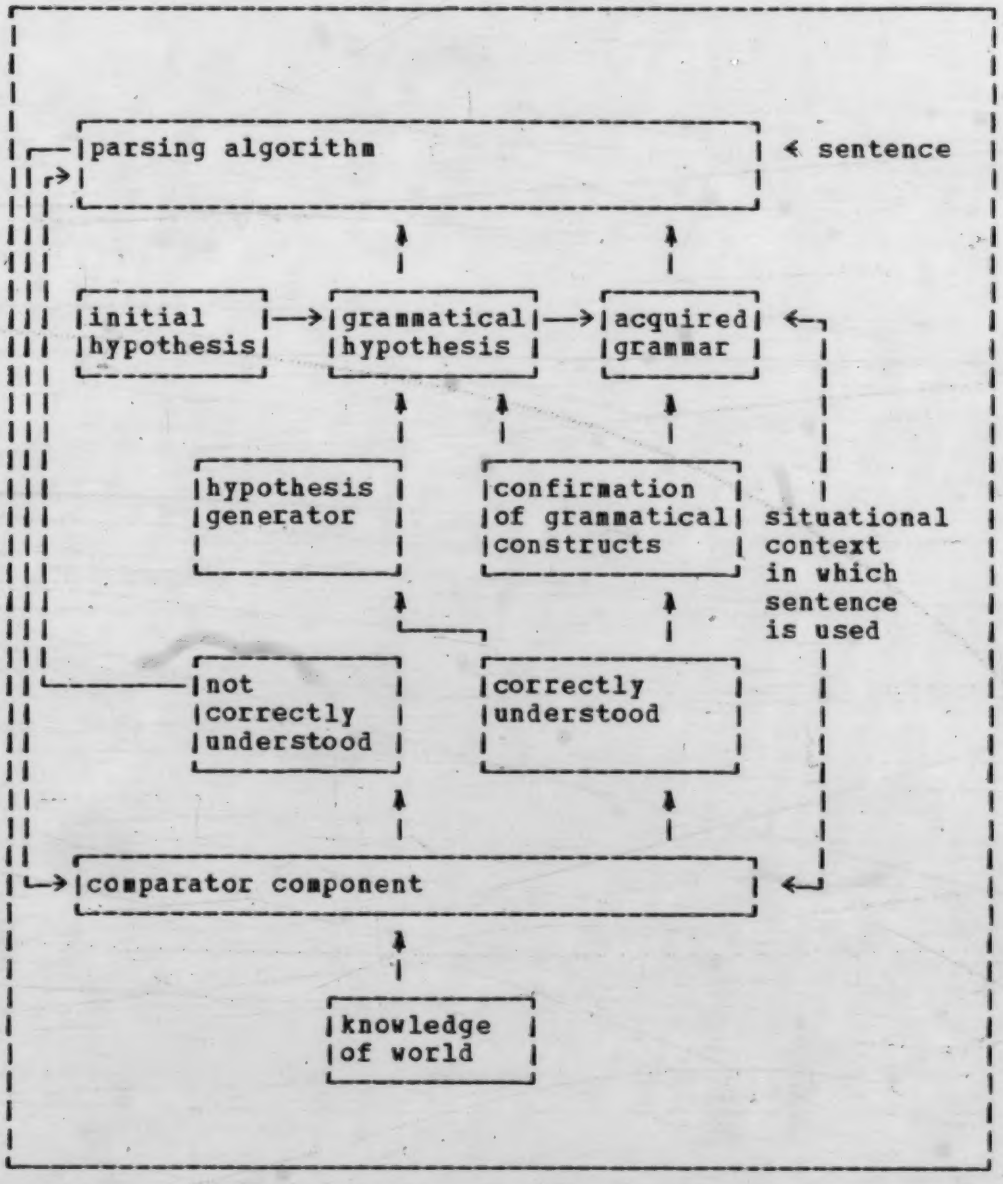


Figure 3.1: Flow Design of Kelley's Language Acquisition Model
<Kelley, 1967, p. 92>

Stage III introduces a new functional relation: subject-of-sentence.

Each sentence which is produced is input to the parsing algorithm, which uses a context-free grammar, consisting of some acquired and some hypothetical rules, and gives either a partial or complete analysis of the sentence. Kelley reasons that children also skip over non-understood parts of sentences and attempt to understand the rest. Concurrent with each stage, the system generates hypotheses to be tested against the input. From the string and the current hypothesis, the parsing algorithm produces a labelled bracketing with identification of the appropriate functional relations.

These analyses are given to the comparator component in an order depending on the amount of the sentence understood. This component determines whether an analysis is consistent with the knowledge of the world. If it is not, the parse is discarded. Otherwise, it is correctly understood and confirmation of grammatical constructs and hypotheses used is incremented. Possibly other hypotheses may be generated as a result. Later, when the model gains more confidence, it may question the validity of the input rather than its own processing.

At the psychological level, Kelley claims, this component matches hypothesized interpretations to the knowledge of the world and the situational context.

However, at the computational level, because there is no "world" available to the system, the structural description is compared to the correct parse formed when the sentence was generated. This assumption greatly weakens the realism of Kelley's system. A child's knowledge of the world bears little, if any, resemblance to a parse of a sentence identifying the grammatical function of each word. The built-in knowledge of the world unrealistically imposes syntactic categories which may be nonexistent in the absence of language.

There is a maturing effect in Kelley's model, but the effects seem to be predetermined rather than allowed to result from changes which are built into the language of the previous stage. A stage should be dependent only upon conceptual growth and prior linguistic ability. In Kelley's model, each stage is associated with a specific hypothesis. Even if one were to give credibility to hypothesis-testing as a legitimate factor in human language acquisition, it is perhaps somewhat hasty to pre-determine the content and the order of these postulates.

In terms of a model, Kelley's system incorporates the environment, which is a necessary component. But he fails to implement a reasonable facsimile of one. His idea of partial understanding of a sentence providing useful data is quite reasonable and worthy of inclusion in any model

definition.

3.1.2 A General Problem Solving Model Applied to Natural Language Acquisition

An adaptive problem-solving model proposed by Harris <1972> employs an objective function as success criterion, an adaptive routine which generates strategies, and strategy testing routines. The objective function rates strategies on their ability to perform some desired task by using them in test situations.

Harris programmed his problem-solver using language acquisition as an example. As the robot moves, the teacher describes the action. The input is a sort of baby talk skeleton, consisting of those words which the robot can map onto concepts of its own mental and physical capabilities. If more than one word is needed to represent an idiomatic concept, the word group is connected by underline characters.

In Phase I, Harris inputs a list of words and a list of their respective concepts. A table of cross-correlations determines the probability that a word matches a particular concept. If the jth concept and the ith word are part of the action and input, respectively, the correlation is

increased. If either the word or concept, but not both, appears, the correlation is decreased. The function Harris uses to determine correlation is based on the previous correlation, a bias (z) equal to + or -1, and m, which is related to the iteration, n, as follows:

$$m = \begin{cases} 16, & n < 32 \\ 32, & 32 \leq n < 64 \\ 64, & 64 \leq n < 128 \\ 128, & 128 \leq n < 256 \\ 256, & 256 \leq n \end{cases}$$

Harris finds the following formula, adapted from Samuel's checker playing program, a useful correlation measure:

$$c(n+1) = c(n)(1 - 1/m) + 1/mz$$

In Phase II, English sentences are mapped onto parts of speech. The strategy is to produce a transformational, context-free grammar with the aid of operators which suggest good grammars.

Harris does not meet the first criterion for a language model (set forth in section 2.4) because his input is unrealistic. During Phase I, the input word-concept pair are valid only on the assumption that the child concentrates on the correct concept in the environment, which one cannot assume. For instance, a child would probably not be able to correctly associate the word "paw" when an adult indicates a cat to him. Even the word "cat" would not necessarily refer to only a cat nor to the entire cat.

Another unrealistic aspect of the input is over-simplified sentences, which seldom contribute to a child's progressional data <Kelley, 1967>. Underlining multi-word concepts, on the other hand, seems no more unrealistic than segmenting the string into words.

This model has made some progress in equating a concept with something other than an arbitrary internal structure. The robot's physical and mental capabilities are programs as well as concepts, providing the environment specified in the second criterion. These processes are pre-divided into part of speech classes, which presupposes a linguistic categorical knowledge prior to linguistic experience.

The model's linguistic ability is progressive in that word associations are made before grammar-learning is attempted, although there is no reason the two processes could not co-occur. Beyond this, Harris has not developed progressive strategies, so the third criterion is also unrealized.

Since the goal of Harris's model was to test a general adaptive problem-solving model, it is easy to understand why the language acquisition process is not very close to human acquisition.

3.1.3 Deep Structures as Card Configurations

By stating explicitly that their system is designed for a robot and not as a model of human behavior, Block et. al. <1975> avoid any direct comparisons of this acquisition method with the steps by which a human acquires a first language. Nevertheless, Block provides certain restrictions in his model which are necessary for a realistic learning environment.

The robot is assumed to combine up-to-date features of a mobile automaton and the Stanford/MIT hand-and-eye robots. The environment is a giant chessboard room about which the teacher and robot converse via Teletype. The robot's learning of syntax is dependent on four conditions:

> 1. An environment to provide something to converse about.
>
> 2. A linguistically competent teacher to provide linguistic input about the environment and to accept or reject the robot's trial utterances.
>
> 3. The robot must have some lexemes associated with concepts.
>
> 4. The robot must be able to learn that the co-occurrence of lexemes can express a relation between the concepts they represent <Block, et. al., 1975, p. 579>.

The robot, with no pre-programmed linguistic information, but with an environment and a linguistically competent teacher providing feedback, can theoretically

acquire considerable linguistic competence. The three types of information it must learn are lexical, syntactic, and pragmatic. Some lexical knowledge must precede syntax acquisition.[1] Pragmatic information is basically non-linguistic.

Lexical information consists of connecticns between lexicons and entities, attributes, relationships, or actions. The strength or weakness of these connections can most efficiently result from teacher feedback. This, as we discussed earlier, is unnecessary in children, although it may be helpful. Certainly, some form of input is required from some competent language user.

The system consists of 4 components: a World Map, which is a four-dimensional representation of a chess-board environment; the Associator, which combines, gates, and passes on information from other components; the Dictionary, which consists of storage bins connected to a teletype and the Associator; and a Syntax Crystal, which results from the learning algorithm. This component allows the recognition of constituent and dependency relationships, distinguishes

---

[1] "One must be able to at least partly understand a string independent of [syntactic] characteristics to recognize that it is a string. Only then can the syntactic characteristics of such a string be acquired." <Block, et. al., 1975, p. 579>

the correct form in parsings of syntactically ambiguous sentences, and portrays the similarities between the constituent relations of paraphrases.

Lexemes with purely syntactic roles cannot be applied to perceptual or World Map data, but rather to relationships among other lexemes. To simplify the process, Block chose not to classify such relations, but to acknowledge only their existence. The Syntax Crystal is the mechanism by which the robot learns to use relational information.

To produce a sentence, the robot selects the morphemes associated with the meaning it wishes to express, and attaches structure cards. To parse a sentence, the lexemes are placed in the order they appear, while the structure cards are added to yield the conceptual relations.

The robot first associates two lexical concepts which co-occur by using two connector cards placed above two non-connected co-occurring lexemes. More cards are added to basic structures by considering the basic structure as a unit to be connected similarly to a third. Though it is not clear how it is accomplished, the authors state that, in the case of "big robot goes", the robot will connect big to robot, then connect this structure to goes. The rectangular cards must remain all in the same direction for making connections.

If substitutions are acceptable, the new lexeme is given the same edge codes as the replaced one; otherwise, a new top code must be generated. The building of additional codes and opticnal codes allow the development of syntactic categories.

The instructor feedback appears to be a mere expedient to the learning algorithm. On the other hand, the necessity of a linguistically competent teacher to the syntax acquisition would imply that the model is not realistic. A child would not receive syntactic correction under normal circumstances, and seldom would have a "linguistically competent" teacher. Also, by taking into consideration only English, the authors run the risk of imposing non-universality on their model.

This apprcach does attempt to operate without pre-programmed syntactic categories, to develop a basic lexicon as a prerequisite to syntax acquisition, and to learn through environmental interaction the meaning attached to the corresponding language. All of these aspects are commendable, and all are incorporated in CLAP, which is described in section 3.2.

3.1.4 Another Problem-Solving Robot

Based upon the Newell, Shaw, and Simon GPS (General Problem Solver), Reeker <1974> proposes a theory of syntactic acquisition. Other than providing an excellent review of the state of the art, the contribution of this work is difficult to assess, especially in light of Harris' prior research. The language is oriented around existing linguistic theory, as are the ideas and diagrams of syntax. It was difficult to find some common ground for comparison with other acquisition systems reviewed.

The proposed PST (Problem-Solving Theory) assumes the goal to be competence, and imitation to be a major motivation. Quite significant is the idea that perception provides reinforcement for the language behavior. The theory is interesting, but means of implementation do not go beyond GPS as it exists.

The work, in general, was of little relevance, since our effort centers around the initial stages of acquisition, which are omitted entirely by a model which begins at syntactic acquisition.

3.2 CLAP

McMaster's <1975> comprehensive language model receives orthographic input from a terminal, as well as input from an

environment. If the environment is a CRT screen, it can be pointed to, windowed, moved, and transformed by either a human or CLAP. No correct parses are input. Utterances are not necessarily simplified to match CLAP's progress.

3.2.1 Components of the System

Linguistic input to CLAP <see Figure 3.2> is in the form of natural adult speech (u), not geared to CLAP's acquisition stage. The non-linguistic input (r,a,s) is natural, within the limits of the environment, rather than artificially constrained for easier learning. Finally, there is no explicit feedback except, during the latter strategies, approval or disapproval and non-linguistic input. At an even later time, linguistic responses will facilitate additional learning.

Both the Parser and the Responder are control structures which may be described as Augmented Transition Networks (ATN's).[2] Each arc of the ATN is labelled with an input, a process, and a weight. Each arc's condition is

---

[2] "A recursive transition network is a directed graph with labeled states and arcs, a distinguished state called the start state, and a distinguished set of states called final states." <Woods, 1970, p. 591>. The network is augmented to provide an arbitrary condition on each arc to be met before the arc is followed.

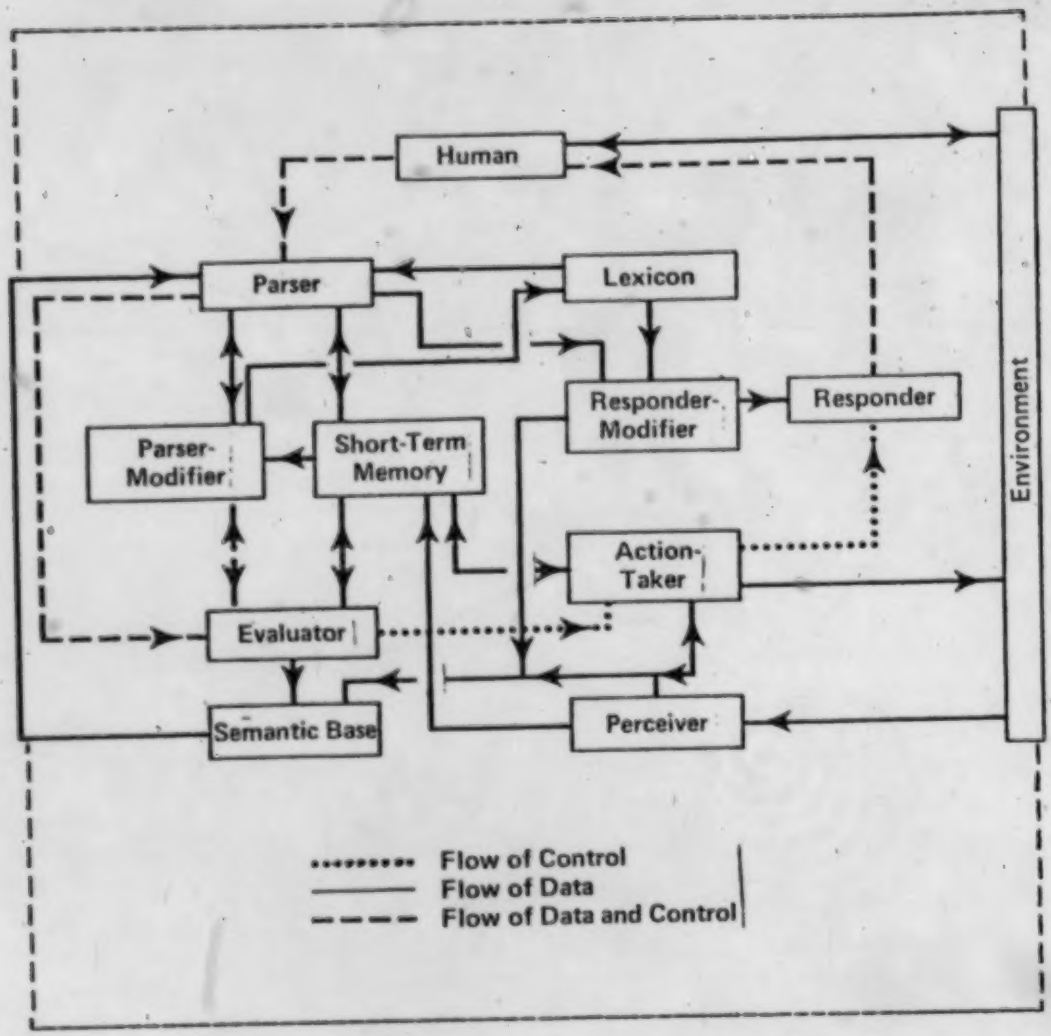


Figure 3.2: Components of CLAP

weighted so that the most highly weighted arc emanating from a state will be chosen. The arc's associated process is executed and the state at the end of the arc becomes the next state of the component.

An input on a Parser arc may be a segment, a branch to a sub-ATN, or a set of conditions on the attributes of the

meaning of the current input segment. This process is a semantic structure or frame. The Parser uses the Lexicon and the set of Focal Structures from Short Term Memory (hereafter referred to as STM) to segment incoming utterances and produce parses. The Evaluator takes this Parse and assesses its credibility. From this evaluation, the Parser-Modifier uses the Parse and Focal Structures in STM to change the Parser's weights, which are conditions on the arcs.

The Responder attempts to construct an output string from the Intention of the Action-Taker. This utterance is an ordered set of lexical items representing the Intention. The Responder's arcs are labelled with 1) a pointer to an element of the Semantic Base, 2) a pointer to a sub-ATN which references an element in the Semantic Base (which may be thought of as inputs), or 3) a pointer to a lexical item or items (which may be null) as an output <McMaster, 1975, p. 126, p. 128>.

The Responder-Modifier changes the Responder in much the same way the Parser-Modifier alters the Parser. This component uses both the Parser and the Lexicon to add to the structures in the Responder and to modify its weights. It examines the Lexicon and, for each concept in the Semantic Base, it attaches the lexical item for which it is a clear-cut meaning. The weight is kept close to the segment

concept weight.

The Responder-Modifier also takes the inputs and processes which label the Parser arcs and translates them into semantic conditions and graphemic outputs labelling arcs on the Responder. The semantic condition is a pattern-matching routine which tries to find the structure of the Intention. The output is zero or more segments.

Many details for this process are missing, including algorithms for traversing the Parser and selecting arcs, but McMaster does enumerate the possible input-to-output translations.

McMaster believes the child neither makes nor tests syntactic hypotheses <McMaster, 1975, p. 134>, but instead attempts to create more complex rules for deriving sense from an utterance. Generalization occurs when the conditions of the rule can be met by a new utterance as well.

The Action-Taker can change the environment, add information to the Semantic Base, or respond to the Human's input by producing the Intention, from which the Responder attempts to build an output string.

The Perceiver, as the name implies, registers environmental changes and stores events in STM. It can also

change the Semantic Base if necessary.

The suggested implementation is a three-dimensional CRT representation which should be rich enough to induce many processes of language acquisition. The more complex the environment is, the greater the concern with whether to explicitly represent relations in the Semantic Base or to generate them as part of the Focal Structure. Winograd's BLOCKS world explicitly represents some, while others are generated in goal-seeking. For CLAP, relations must be explicitly represented, at least in the Focal Structure.

The Semantic Base is the essential interface between the Human and CLAP. It is the BLOCKS world, structurally represented in a uniform way for actions, inferences, events, etc., allowing uniform learning procedures to be used by the Parser-Modifier. The STM contains an event list, which includes information as to Parses, Responses, Foci, and Human non-linguistic input at the occurrence of each.

3.2.2 The Learning Strategies

Rather than emphasizing production, McMaster chooses to consider the <u>strategies</u> which are applied to comprehension aspects of the acquisition problem. The building of the

Parser is the governing function of CLAP's acquisition process. Production is simply a result of structures built in the Parser <McMaster, et. al., 1976>. CLAP does not have the complication of cognitive development found in children, so the strategies are admittedly an oversimplification of human acquisition strategies. During the first strategies, word-concept associations develop as links which have weights indicating the probability that the link is correct. McMaster explains the first two strategies in some detail, while the last three are left relatively nonspecific <King, et. al., 1976>.

3.2.2.1 Strategy 1: Segmentation and Meaning Association

Strategy 1 involves the establishment of weighted links between the lexical items and the Semantic Base. At first, lexical items may include many incorrectly-segmented morphemes which eventually atrophy out of existence. Weights between these segments and the concepts help determine the most likely parse.

As a child learns to recognize repeated chunks of the input stream through the recognition of morphophonemic boundaries, CLAP must learn to recognize morphographemic bounds. Parsing in Strategy 1 consists of breaking the input into meaningful segments <See Figure 3.3>. At first

the input is segmented into single characters by that part of the Parser called the Segmenter. The Parser-Modifier then creates a new segment from each pair of adjacent segments. The Parser then attaches pointers from each of these segments to each concept associated with the focal region. Depending on the sophistication of the Semantic Base, these may include concrete referents, classes and relations, cognitive competence, and the ability of the system to affect the world.

The establishment of segment-concept links is used for generating future parses. The Evaluator compares the interpretation, based on concepts in the focus, with those concepts closely linked to the input, and passes its results to the Parser-Modifier. The weight will be lowered if the focus does not include a particular concept. Otherwise, association proceeds as before. Again, the Parser-Modifier attempts to join adjacent segments to add to the lexicon.

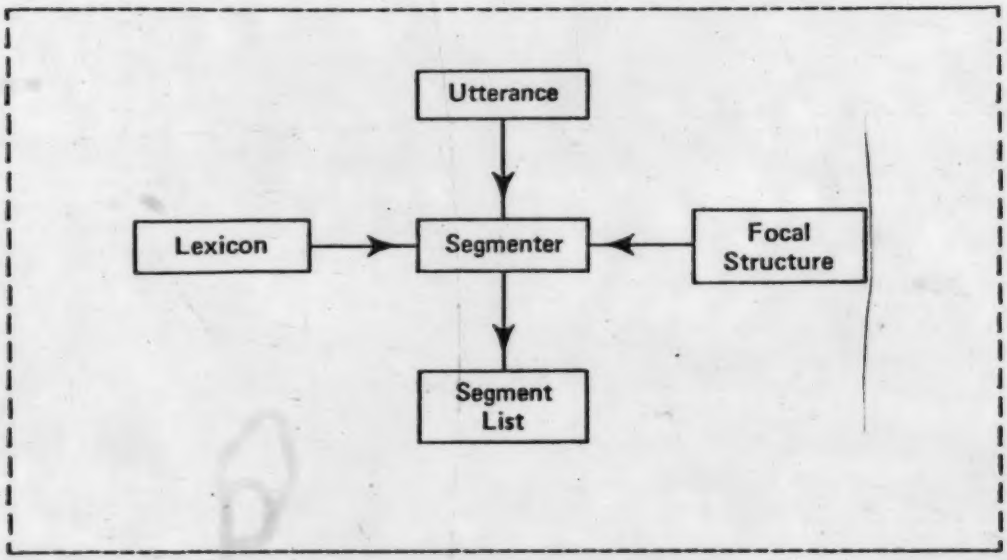


Figure 3.3: Parsing in Strategy 1
<McMaster, 1975, p. 98>

The Action-Taker's activity depends on the input. CLAP's global goals are to accumulate knowledge, to communicate its knowledge to the human, and to receive approval. Therefore, CLAP looks for something it can add to its Semantic Base, something indicating a desire for CLAP's knowledge, or something indicating approval. If it recognizes, in the list of concepts expressed in the input utterance, any completely specified procedure, the Action-Taker may evoke that procedure. During Strategy 1, it cannot add to the Semantic Base because the concepts are not specifically related. Likewise, it cannot yet respond.

If an action is input, the Action-Taker can select part of its conceptual structure, attempting to respond by

outputting a list of those concepts it can map onto lexical items. Here lies the correspondence with a child's first one-word utterances.

In the case of disapproval, the Action-Taker looks to its previous output, stored in STM. It can diminish the weights on links between words and concepts, depending on its confidence in those links.

If a sentence is input, the Action-Taker responds as in the case of an action, except that it must use some heuristic to determine the Focal Structure.

During Strategy 1, then, CLAP associates concepts with input strings while learning to distinguish, within the linguistic context, those strings with associated concepts.

3.2.2.2 Strategy 2: Linear Ordering

> As soon as the Lexicon has developed to the stage where CLAP can understand more than one word (segment) in an incoming utterance it can begin to attach import to word order. However,...Strategy 2 is really concerned with building structures from a few morphemes. <McMaster, 1975, p. 104>.

The building blocks of the future Parser are weighted segment-concept links established during Strategy 1, and the Structure Builder, an ATN built by Strategy 2. This is the

first time CLAP begins to use an ATN.

The Segmenter continues as in Strategy 1. The Concentrator examines the Lexicon, Segment List, and Focal Structure to produce a Target Structure, the smallest and most highly weighted semantic structure which both appears in the Focal Structure and contains all concepts associated with segments in the Segment List. The Parser-Modifier builds the first of the future Parser by examining the Target Structure and building a network which accepts the Segment List and outputs the Target Structure. Each arc in the network is weighted so as to indicate the subjective probability that it is the correct arc to follow in a particular situation to produce the desired Target Structure.

With natural input, there will always be some words which have no direct referent in the environment, yet change the meaning of an utterance, their own meaning lying within the relationships of the referents. McMaster suggests ways of handling these morphemes during this strategy, but this will not be discussed here.

3.2.2.3 Strategy 3: Structural Generalization

Generalization is required to make the parser efficient

and to generate novel utterances. It also helps organize the parser. Similar semantic and syntactic processes which label the nodes of the Parser can be combined into a single label and process. The old arcs, if not used, will age their way out of the Parser. Likewise, faulty generalizations, lacking confirmation, would eventually be discarded.

McMaster defines three types of generalization which might be useful in implementing Strategy 3. Generalizations are based on the structure of the Parser, and directed by semantic regularities and processes which label the arcs. These are only suggested guidelines and are not a strict specification of this strategy.

Semantic generalization (Strategy 3.1) involves the regularity of semantic characteristics on the arc labels. For instance, if two or more processes on two or more arcs require filling a slot with distinct concepts which share attributes or values at some level, new arcs are created to parallel the old ones, which generalize the process. The new inputs resemble feature-matching procedures.

For instance, suppose two arcs have the inputs I(1) = B6 and I(2) = p1, and corresponding processes P(1) = "insert :B6 in *3" and P(2) = "insert :P1 in *3". (B6 and P1 are lexical items; :B6 and :P1 are concepts; and *3 is a slot in

the predication). Since both :P1 and :B6 belong to the class of entities called #THING, this is the condition placed on the new, generalized arc, which must be met before the arc can be followed.

Strategy 3.2 is what McMaster calls syntactic generalization. This reorganization of the Parser is on the basis of topologically similar structures occurring in sequence. In the case where two arcs have similar inputs and the same procedure, recursion is introduced. This allows such occurrences as an unlimited number of modifiers. Strategy 3.2 would necessarily follow the semantic generalization of Strategy 3.1, because there would seldom be instances of the same lexical item appearing twice in succession.

Strategy 3.3 generalizes congruent parts of an ATN, causing the new arc inputs to be a sub-ATN which generalizes the previous structure-building operations. An instance of this would be the occurrence of a prepositional phrase in various parts of the sentence.

3.2.2.4 Strategy 4: Conflict Resolution

Occasionally, discrepancies between the interpretation of the utterance and the environment cause noisy input to

the system. Eventually, problems of interpreting such conflicts as lying, hypothesizing, making negative statements which are true about the environment, requesting information in a statement form, and commands to change the current situation, must all be linguistically mastered or they will confuse the system.

Even a child will be confused when presented with a large number of untruths. This must be avoided in CLAP. To overcome the other difficulties, the Evaluator can, by some heuristic, either build a hypothetical world, negate the parse, hand the parse to the Action Taker, or, if there is no direct contradiction, adjust the Semantic Base to agree with the parse.

3.2.2.5 Strategy 5: Using Discourse

McMaster has very little to say on this subject. It is not clearly understood how humans employ such complicated processes as anaphora, quantification, hypothetical worlds, and referential ambiguity in daily discourse.

The usual area of concentration in mechanical language systems is the sentence. Wilks <1973a> enlarges his context to the paragraph, but humans can connect ideas far beyond that limit.

3.2.2.5 Strategy 5: Using Discourse

This strategy requires the parse to supplement or replace the Focus. In responding, CLAP's Responder accepts the Intention of the Action Taker and attempts to produce an utterance.

3.2.3 Evaluation

CLAP represents a step towards the realization of a language acquisition system which is both realistic (the major goal set forth in Chapter 2) and programmable. The major drawback to using CLAP as a model is the lack of necessary detail for implementation, especially in the later Strategies.

The fact that CLAP uses graphic input seems analogous enough to the input of speech, fulfilling the requirement of realistic input. Although graphic input lacks cues often provided by vocal inflections and stress, it has not been established how important these factors are in the learning of segmentation, because languages vary in the extent to which such cues convey meaningful hints. An advantage of the graphic input is the limited number of primitive symbols an alphabet has, as opposed to the large number of phones in a human dialect.

An environment also exists for CLAP. Throughout,

McMaster suggests implementation using a SHURDLU-type BLOCKs world <Winograd, 1972> on a CRT (see figure 4.1). If the equipment were available, a robot would insure a richer environment and the ability for CLAP to conceptualize itself and its processes as part of the environment.

The strategies provide a means of designating CLAP's progress, although there is no cognitive maturing effect beyond this domain. By avoiding the introduction of cognitive development, no unrealistic psychological progression is imposed on the system. The lack of any true cognitive development could be construed as a major weakness of CLAP. On the other hand, there could be disadvantages to implementing a dynamic maturity for CLAP, not only because of the lack of a specific model of the human processes, but because an experimental control would be lost. Assuming one of the goals of a computerized language acquisition system is to learn more about how humans acquire language, it would be advantageous to study this in the absence of maturing effects.

As far as it is specified, CLAP appears to be programmable, though it is not always efficient. The entire segmentation process suggested for CLAP stresses enumeration of all possible segments rather than the generalization noted so often in other aspects of a child's acquisition. It is interesting to consider whether children might go from

the top down, instead, placing sentences or parts thereof, up to a certain length, into the primary lexicon, then breaking these down by pattern-matching routines. For instance, if "thedogisbrown" and "hownowbrowncow" were both in the lexicon, and each pointed to the concept, :brown, among others, a routine could generalize the situation, using both spelling and meaning as clues, to create a new lexical item, "brown", with pointers to all the meanings these two sentences had in common. Faulty generalizations would drop out from lack of use, as would those sentences input as lexical items.

When the meaning of "brown" became established, sentences like "hownowbrowncow" could be interpreted as two items, "hownow+cow" and "brown". In this manner, disconnected morphemes could be lexicalized. The Lexicon itself could be either a separate ATN or part of the Parser. A Lexicon-Modifier could adjust the weights on links.

Segmentation in Strategy 1 is an area in which a significant amount of research is needed before it can be practically implemented.

No doubt, many other problems will arise as CLAP's Strategies are implemented, but it is hoped that the overall approach will survive the test of applicability.

3.2.4 A Test of Plausibility

A Vocabulary Acquisition System (VAS) attempts to demonstrate, by carrying out one of the first tasks of CLAP, the feasibility of implementing CLAP. VAS attaches the meanings of concepts in a simple CRT-like BLOCKS world to descriptive words. VAS does not implement all of Strategy 1.

The VAS environment is a simple CRT-like display without the CRT. In other words, although size and position are indicated within the semantic base, there is no implementation of a display.

The internal representation is a predicate calculus notation similar to Winograd's<1972>. Each predication provides a potential insertion pattern to be an arc label on the ATN. One problem immediately following from this is the awkward manner in which the semantic base must be scanned. For instance, to display an object, :B2, which is a red cube with side 50 and located at point (100 200 200), one would need the following information or some similar representation:

```
(#AT :B2 (100 200 200))
(#SHAPE :B2 #RECTANGULAR)
(#COLOR :B2 #RED)
(#SIZE :B2 (50 50 50))
```

In Winograd's and in McMaster's representation, one would

need to scan each list in the semantic base for those whose second element is ":B2".

A second criticism of this representation is that (#IS a b), which indicates that a belongs to the class b, superimposes a classification scheme to which a child might not have access until his language delimits such categories.

The linguistic representation, too, follows Winograd's format. Associated with each lexical entry is the indicator WORD with value (WORD) and the indicator SMNTC with the value

```
(WORD ((MEANS ((<concept><weight>)
               (<concept><weight>)
                        .
                        .
                        .
               (<concept><weight>) )) ))
```

WORD can later be the part of speech, once CLAP progresses to the point where it can differentiate them. The weights here are simply counters of the simultaneous co-occurrences of the word and the concept.

Sentences are input with complete punctuation and are pre-segmented. VAS learns no segmentation. A child's verbal input includes pauses, intonation, and gestures. The first two roughly correspond to punctuation. VAS considers only gestures like pointing in the process of determining a focus. Ideally, a sentence and a focal point in the two-

dimensional representation of the BLOCKS world would be input. In VAS, though, only a list of objects constitutes the preliminary focus. An initial lexicon is input with no meanings, as there is no management scheme to delete lexical items which have no environmental referent.

"There are three classes of predicates in the BLOCKS world:

1. One-place attributive predicates. For example, (MANIP x) attributes the property of manipulability to each concept in the list x.

2. Two-place attributive predicate. For example, (#IS x y) attributes membership in the set represented by the concept y to each of the concepts in the list x.

3. Relational predicate. For example, (#SUPPORT x y) means that the non-commutative and transitive relation of "supporting" exists between the object x and each object in the list y.

Each concept c in f is examined, and each class of predicate is processed as follows:

1. For each Class 1 predicate p in which c occurs, p is added to the focus.

2. For Class 2 predicate p in which c appears as a first argument, each concept in the second argument is added to the focus.

3. For each Class 3 predicate p in which c occurs as the first argument and one of the concepts in the second argument also appears in the focus, p is added to the focus."
<McMaster, 1975, pp. 144-145>.

Word-concept co-occurrences are counted, the result

being stored in an i by j matrix. The number of co-occurrences of the i'th word and the j'th concept is the (i,j)th entry in the matrix. The total number of times a concept c(j) has occurred, the number of occurrences of word w(i), and the ijth entry are used to derive a weight between the two which, assuming a larger value of the co-occurence:

1. Is small if both w(i) and c(j) have occurred often.

2. Is also small if w(i) is seldom used while c(j) is used frequently.

3. Is larger if w(i) occurs frequently and c(j) does not.

4. Is largest when both w(i) and c(j) have occurred infrequently.

If u(c) equals the number of times a concept c has appeared in a focus, u(w) equals the number of times a word w has been presented in the input, and u(c,w) is the number of times a word w has appeared in the utterance at the same time that the concept c appeared in the corresponding focus, we can construct a function to derive the correlation between the word and concept as follows:

$$P(u(c), u(w), u(c,w)) = u(c,w)\ (2 - u(c)\,m/u(w))$$

where m, whose value was chosen through experimentation, is at present .21, with no claim to being the optimal value for all corpora. In fact, the function of m is not understood; it would be interesting to lock into this problem.

The results of the implementation are impressive. From

a noisy input of 219 utterance/focus pairs, VAS successfully learned 9 out of 16 direct associations. With a corpus more oriented towards helping VAS learn, 18 out of 24 associations were learned with 39 (u,f) input pairs.

VAS does not learn to differentiate between two concepts which always co-occur. Children also produce such errors. VAS can also wrongly associate lexical items in the corpus whose proper concepts never appear in the focus. Such instances were due to the use of Misleading Extraneous Words, and was the fault of the noisy corpus. Finally, some concepts occurred so often that, when the correct word did occur, the Excessive Concept Usage (ECU) lowered the value of the correlation function to the extent that the correct concept was not chosen as the meaning of the word. McMaster suggests that this may be corrected by incrementing u(c,w) for each time the concept appears in the focus. Another solution to ECU might be to alter the correlation function.

4

A VERBAL ACQUISITION MODULE

In spite of its shortcomings, CLAP provides the most complete framework proposed to date for an acquisition model. Until the proposed concepts have been applied, we can only speculate as to their effectiveness in the task of language acquisition. VAS made a start in the direction of implementation. It is hoped that VAM will further the completion of the first strategy, as well as reveal areas which could be reformulated.

The Verbal Acquisition Module is an attempt to overcome some of the shortcomings of VAS. Schank's conceptual representation will be adapted to the module, altered because of the limitations of the environment and needs of the module. Much of VAM's capability and methodology is identical to that of VAS. Therefore, we will emphasize the changes which have been made to accommodate acquisition of verbs in the model.

Movement in VAM's environment enriches the conceptual vocabulary beyond that of VAS. VAM either observes an action, performs an action, or, like VAS, receives as input a pointer to one or more objects. When it perceives an utterance, it compares the expanded focus (see below) with the words of the utterance, and computes a correlation between each word and concept of the inputs. The calculation of this correlation will be discussed in section 4.4.

4.1 Conceptual Structures for VAM

VAM's environment is like a CRT display of BLOCKs, as in VAS's world. Although the display is not implemented, VAM's data structures contain all the necessary information for such a display. There is no gravity, mass, or temperature in VAM's world, although there are three dimensions and a representation for color.

Since there is no CRT display to be used in pointing, windowing, and moving objects, and as the Human is imperceptible to VAM, there is no viable means of distinguishing between an action performed by VAM and one performed by the Human. The only distinction is between movements which name VAM as the agent, and watching objects change positions, as a result of action by the Human. The

only purpose of this convention is to see if VAM could learn its name in relation to the concept of itself. Because this is a superficial imposition on the system, in that VAM cannot make decisions to act upon the environment, we did not expect impressive results in this aspect, but were somewhat surprised, as we shall discuss later.

There are nine physical objects in the environment, each represented by a list in the semantic base, and each having its own name. Three are cubes, two are rectangular blocks, three are pyramids, and one is a box, capable of containing other objects. In addition, there is VAM's arm, a black line-shaped figure which hangs down from the top of the scene (see figure 4.1).

As mentioned in Chapter 3, the conceptual framework of Roger Schank provides a springboard for the development of VAM's data structure <see Figure 2.1>. Schank's descriptions center around events, whose structure would be learnable by VAM only if the STM (short-term memory) were implemented. VAM does not have access to a memory except for the usage counts of words, concepts, and mutual occurrences. Neither is time represented to VAM. The time during which an event occurred could not be represented in our drawings. Time between events was omitted because we were not implementing short term memory in which to store an entire event.

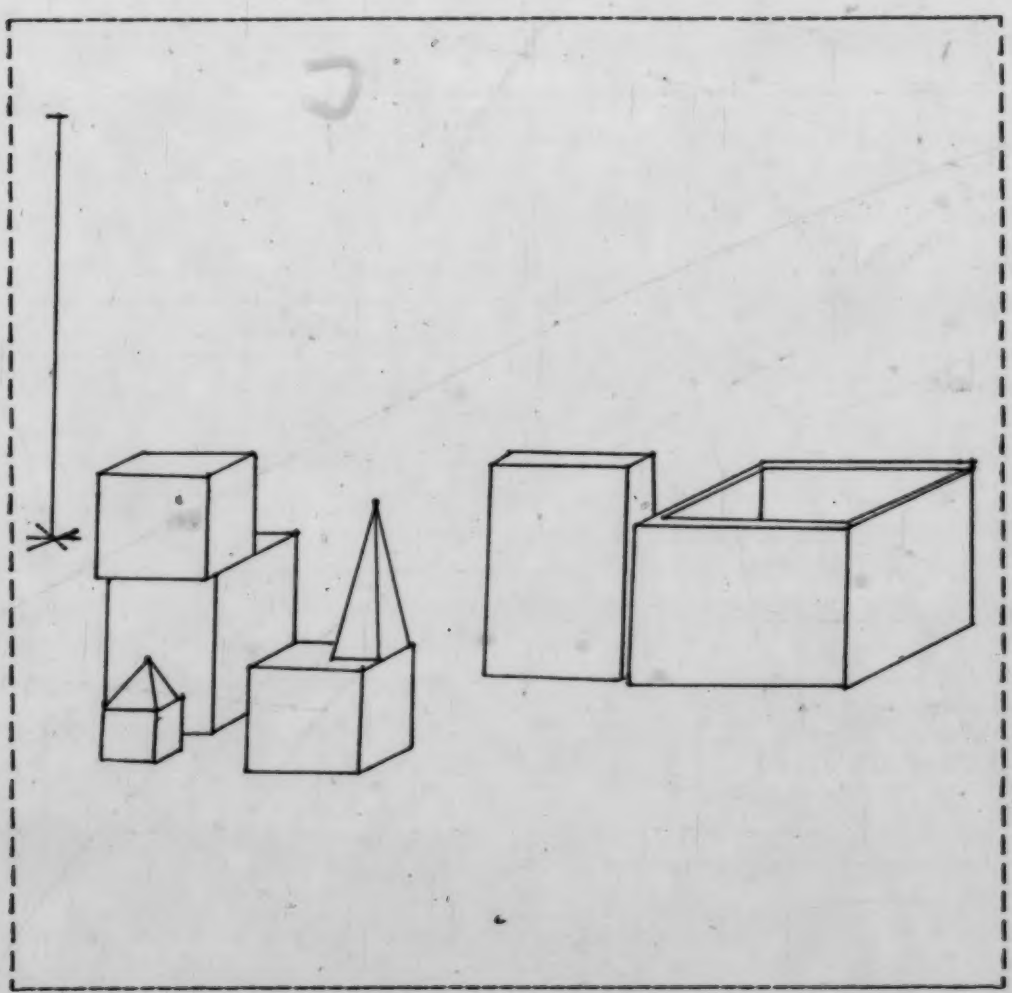


Figure 4.1: The environment for VAM

Instead of the representation of an event as a nodal net structure, we will list in the focus those concepts which would make up an event in VAM's world, including the name of the event itself and the more primitive actions from which it is formed. It is these events, or programs

representing various levels of structure, which actually create the focus, as well as handling the necessary environmental manipulations. We decided to include all levels, including primitives, in an effort to avoid language dependence in our representation. It is assumed that a child at the same stage as VAM would be capable of extracting the concepts of the action from its total occurrence.

It will be recalled that Schank characterizes a conceptualization as centering around an actor-action mutual dependency, or an object-state one:

> A conceptualization consists of either an actor-action-object construction or an object-state construction. If an action is present then the cases of that action are always present. One case of an action is instrumental which is itself a conceptualization. <Schank, 1973b, p. 12>

Our model is concerned with the former construction, the object-state one only performing an instrumental function for some actor-action-object configurations.

4.1.1 Components of the Actor-Action Conceptualization

Since an action cannot occur without an actor, VAM, being the only actor in the environment, is the only one who

can perform an action. Yet VAM has no physical representation of itself other than its arm. VAM's arm, which can be the actor of the instrument of the action, can have attributes, containment and location. (Hereafter, "attributes" will include containment and location, unless otherwise specified.)

The time and location of a conceptualization are not concepts to VAM, because of a lack of memory and the absence of any other locations with which to contrast the location of an event, which always occurs in the same CRT-screen location.

The action itself must be limited to those primitive actions which are relevant to VAM's viewpoint. As we shall describe below, some of these can be broken down into a series of actions, while some have inferences which should be included in the focus. We also believe Schank overlooked some primitives, so we add ACTs which could not be expressed in terms of the fourteen primitive actions.

The ACT could have attributes, but events are registered by VAM as discrete environmental changes, which limit the display of attributes such as speed. Therefore, no attributes of actions will exist in VAM's focus.

Naturally, the object of an action, as well as its attributes, can and must be included among VAM's concepts.

An object will always have its state represented as its data structure at the occurrence of a particular event. This includes an object's attributes of size (three dimensions), whether or not it is visible (interpreted by VAM to be its positive or negative existence), its color, its shape (classification), its location (three co-ordinates), and its contents (a list). Each attribute value occupies the specified position on the list defining the object. For example, :B1 is defined as

(((1 1 1), T, :RED, :BLOCK), (1 0 1), ()).

Color and size are permanent attributes; the rest can be changed.

4.1.2 Cases of the Action

The directive case will be defined as the resulting interrelationships between the object of the action and other objects in the environment. In this instance, although the relationships and the object names will be included in the focus, we have arbitrarily decided that such objects were not important enough to have their attributes included in a uniformly-weighted focus. This would tend to simulate the child's orienting response to what is moving or changing, rather than to the background objects.

Objects in the BLOCKs world may hold relationships with other objects, but it is doubtful that reciprocal concepts could be correctly associated with one word until the later strategies, when word order has become significant. For instance, given the concepts left-of and right-of, VAM could not differentiate them at its present level of experience. That is, VAM would have to know the names of both objects before such contrasting concepts could become associated with appropriate words. Therefore, in VAM's world, relations are defined with a single concept representing both a relation and its reciprocal: BESIDE, FRONT-OF (or behind), SUPPORTS (or is-supported-by), BIGGER (or smaller).

When Schank speaks of the recipient case, he gives examples like catching a thrown ball. There is nothing in VAM's experience which could correspond to this case (as two actors are required here), except possibly what we have already described as the directive case. Whichever we call it, the thing to which another object is transferred will be included in the focus. However, we again arbitrarily decided to limit our focus to the resulting state of affairs, excluding the prior situation. It intuitively makes sense that the result of an action would be more noticed than the prior configuration, especially when the comments regarding an event follow the occurrence.

Finally, the instrumental case is another

conceptualization itself, either of the object-state or the actor-action variety. We will discuss this with examples when we discuss the particular actions and their inferences.

Thus, we are left with the conceptual representation of an event as shown in figure 4.2. (The reader may find it interesting to compare this to Figure 2.4).

4.1.3 Relevant ACTs and Their Inferences

Of Schank's fourteen primitive actions <see section 2.2.1>, only three physical ACTs and one global ACT are applicable to VAM's world. These are PROPEL, MOVE, GRASP, and PTRANS. Mental ACTs all require an understanding of the mental processes within the individual. Schank indicates that young children are unaware of the Mental ACTs, MBUILD and CONC, within themselves. Likewise, we doubt that MTRANS develops very early. Furthermore, the internal functions of VAM's mental processes is a big problem without Short-Term Memory, cognitive development, and proceedures for problem-solving.

Instrumental ACTs would probably not be learnable by a system with no sensory organs. VAM has representations of pseudo- aural and visual input, but these (LISTEN-TO and LOOK-AT), being present in every focus and sentence, would

not develop any clear meaning.

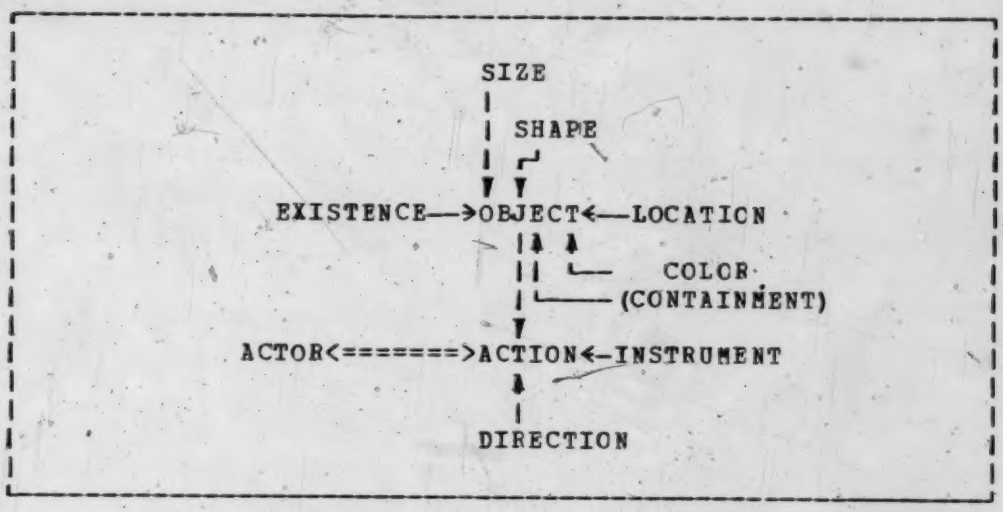


Figure 4.2: Actor-Action Conceptualizations as Seen by VAM

The global ACTs are PTRANS and ATRANS. The latter, which involves possession, is too abstract to represent in VAM's world. PTRANS is the change of location for an object.

Global ACTs are used to represent the way a human focuses on the result of an ACT, rather than on the ACT itself. Although Schank insists that PTRANS must result from a physical ACT upon an object, we use PTRANS to bypass our problem of having HUMAN action on the environment. PTRANSing of that object is what VAM perceives.

Schank <1973b> claims that the physical ACTs are all that one can perform upon an object. None of the fourteen

he lists can portray the concept of turning. Therefore, VAM will understand a broader PTRANS, consisting of four types, three of which define rotation through one of an object's three axes, and one which maintains the original definition: TURN (Y-axis), TWIST (Z-axis), DROP (X-axis), and RELOC.

Schank <1973b> specifies which inferences can be drawn when a particular ACT is present. Many refer to the intentions and feelings of the agent, which dc not concern us here. Of the main inferences of PROPEL, the relevant one is PTRANS (assuming that the object is not fixed). Furthermore, the instrument of PROPEL, unGRASP, can also be inferred. (In contrast to Schank's <1973b> theory, PROPEL is not used as the instrument of PROPEL in VAM's world, as domino reactions are not allowed. Also, MOVE is omitted, as striking an object to propel it would be difficult to demonstrate in drawings in sufficient detail.)

The only inference of PTRANS is the new location of the object, and thus its new relationships with new objects. Its instrument is MOVE or PROPEL, but this is circular when PTRANS may be inferred from MOVE and PROPEL.

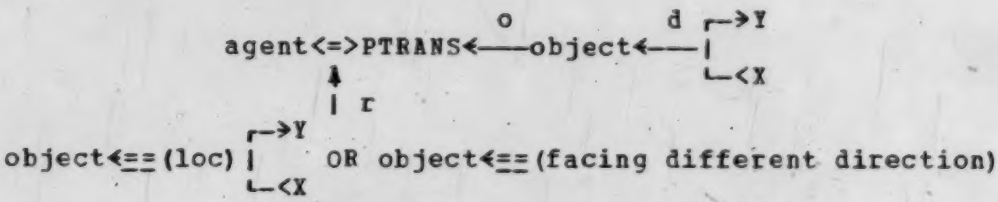
For some reason, Schank states that PTRANS is inferred by GRASP. We cannot imagine such an instance, and Schank gives no examples of any. The instrument of GRASP may be MOVE, but is not necessarily so. Only VAM's arm can move,

since there are no fingers.

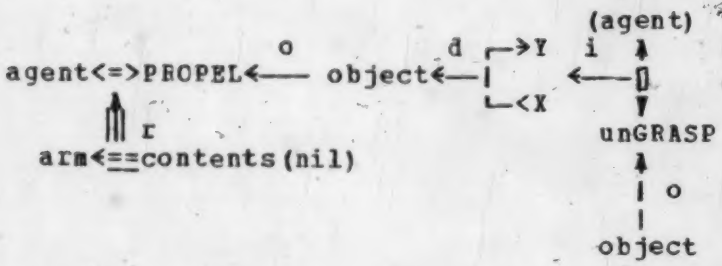
No data was available on the instrument of MOVE. There may not be one. Schank lists no helpful inferences for this ACT, either, but PTRANS should follow if the body part is somehow attached to another object, as in the case of GRASPing something and MOVEing that hand.

To better understand VAM's conceptual capabilities, we will shortly examine (section 4.3) how the higher level events interrelate the ACTs and how foci are drawn from various combinations of these. In summary, the structure of the elementary events for VAM and their conceptual dependency formats, including inferences, are:

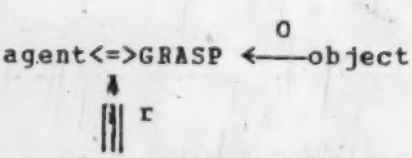
PTRANS: to cause an object to change states.



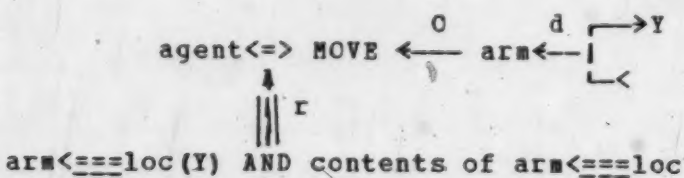
PROPEL: to apply a force to an object, unGRASPing it in the process.



GRASP: to grasp or let go of an object.



MOVE: to relocate a bodypart (specifically, the arm).



4.2 Objects in the Focus

Any time the arm is considered to be a part of VAM's focus of attention, the concept of the arm itself, its color as defined by the pen parameter used to produce a hard copy, and its shape are included in the focal list. We assume its contents to be already part of the focus, as it would be central to any action involving the arm.

The object of the action is central to the static focus. The object itself, its attributes, and its relationships with other objects are included.

Any block can enter into the following relationships with any other block:

BESIDE	One block is touching sides with another.
FRONT-OF	One block touches another front to back.
SUPPORTS	The center of gravity of the block on top is directly over the lower one. There may be other blocks intervening. The box may not enter into this relationship.
BIGGER	The shape of two touching objects is the same, with each dimension of one greater than the corresponding dimension of the other.

Each block may hold one additional relationship to the box:

INBOX	The object lies within the box.

The objects and relationships are merged into the focus in a manner described later.

4.3 Building a Focus

Based on the primitive ACTs, more complicated events can be built, deriving their foci from the structures of the individual ACTs involved. For any primitive action, a single object is specified. The particular program places its own name, object, and agent in the focus, makes required state changes on the object's attribute values, then expands the focus from each object.

1. For each object already in the focus, if it is related to any other object, then that relation and that object are added to the focus.

2. If the box is in the focus, its contents are added to the focus.

3. Finally, the attribute values of all objects in the focus, including the arm, are added to the focus.

As in VAS, a focal object set may be indicated to VAM. In the case of a pointer, only those objects pointed out are placed in the focus, along with their relationships to each other and their attributes. We also decided to place the event's (or program's) name in the focus, as the entire event should be a learnable concept as well as the component parts, even though it may have no direct lexical link in the

English language.

Events were shown to VAM, consisting of one or more primitive ACTs, and their associated cases. In the instance of PROPEL, the only occurrence was in the act of THROWing, which included the original definition of PROPEL in addition to PTRANS, which was implied. Were we to differentiate THROW from PROPEL, VAM would attach identical weights to the two concepts for each word involved. Thus, we used PROPEL as the more complex event. Its focus is:

```
VAM PROPEL k
k PTRANS a(k) r(k) o(k) a(o(k))
VAM GRASP arm a(arm)
```

The human can cause an object to be PTRANSed. Although VAM cannot "see" the human, it would notice the effect on the environment. The name of this program is WATCH. Given that the object, k, has attributes a(k), and that it is in relations r(k) to some other objects o(k), the focus for this event would be WATCH and the focus of PTRANS:

```
WATCH PTRANS (TWIST, DROP, TURN, or RELOC) k a(k)
r(k) o(k) a(o(k))
```

Although k serves two functions, as both agent and object, it is related to its environment only once. This program would move k to its new position before forming the focus.

VAM can MOVE its arm, possibly PTRANSing its contents, k. Using squared brackets to indicate conditional

inclusion, the focus for movearm would be:

```
MOVEARM VAM MOVE arm a(arm)
[ VAM PTRANS k a(k) r(k), o(k), a(o(k)) ]
```

The second part is included if the arm contains the object k. VAM, of course, would appear only once.

VAM can GOGET an object, k, which requires GRASPing k, possibly necessitating a preceding unGRASPing of former contents, k', and/or MOVEing towards k. In accord with our decision to de-emphasize the pre-action, k' will not be completely expanded in the focus in its relationships to other objects. The focus would be:

```
GOGET VAM GRASP k a(k) r(k) o(k) a(o(k)) arm
a(arm)

[ VAM GRASP k' a(k') ]
[ VAM MOVE arm ]
```

If VAM lets an object go, PTRANS may be implied unless the object is already resting on something. The focus for letgo is:

```
LETGO VAM GRASP k a(k) arm a(arm)
[ k PTRANS k a(k) r(k) o(k) a(o(k)) ]
```

For the experimental data, PTRANS will always be included in letgo, because a simple unGRASPing action was too subtle for appropriate data to be collected.

Finally, VAM may transfer an object, or turn it, by

MOVEing an arm, PTRANSing a GRASPed object. The focus for this is:

```
TRANSFER VAM PTRANS k a(k) r(k) o(k) a(o(k))
VAM MOVE arm a(arm)
```

The grasping of the object, if necessary, would be part of the pre-action, and thus is omitted.

4.4 Evaluating the Associations

Each lexical item has associated with it a usage counter and a list of co-occurring concepts with a count of the number of times the word and concept have appeared simultaneously.

A numeric value is used to determine the weight of a concept-word correspondence. Harris proposed a function based on time and the weight at time t-1. This involves keeping records of the previous correlation, which could be done. However, the use of time is an artificial imposition upon an environment in which time is not expressed as a concept. Because of this drawback, and because using McMaster's formula would provide more controls on the comparison of VAM to VAS, we choose to correlate concepts and words as follows. The program keeps a record of

1. The number of times a lexical item has occurred,

2. the number of times a concept has appeared in foci,

and 3. the number of times any concept and any word have co-occurred.

With this information, VAM is able to decide the most likely meaning for a given word.

Experimental material for VAM was programmed as a series of 31 random events, whose resulting environmental configurations were plotted by computer use of the data structures. To generate the first of VAM's two experimental corpora, the resulting series of plots was then shown to a young child's mother, who provided the corresponding input data by describing the events to her 1-1/2 year old daughter. It was hoped that this would provide input similar to that which a child receives. The following is a representative sample of corpus 1:

> Now the blue box is turned sideways so the thinner part is facing towards the green box which is in front of it.
>
> Looks like a book.
>
> And now we've put the small red box on top of the green box.
>
> Now the little green triangle is on top of the red box which is in front of the rectangle that looks like a book.

VAM's second corpus was constructed by VAM's designer, with

the hope of observing the system's optimal performance.

At random points in the experimental runs, we ran a vocabulary test on VAM to determine a word's meaning to it (or the closest collection of meanings). If the word was associated with a concept which is intuitively acceptable, the word was considered to be learned. If the meaning was not clear-cut, but if all the meanings associated with the word together comprised a general definition, the word was considered correct.

Not only do the number of words learned interest us, but some particular associations are quite unexpected, and will be described below.

4.5 Preliminary Experiments and Results

The experiments with VAM were aimed primarily at noting trends, rather than proving particular hypotheses. We also wanted to see which of several pairs of alternatives tended to produce the best results. For instance, VAS's lexicon was pre-defined by the program, and consisted of those concepts which McMaster deemed learnable. Words such as the and if were eliminated as functional words which could have no meaningful link to an environmental concept. It was assumed that these would atrophy out of the lexicon when no

clear-cut meaning evolved.

Perhaps, this could eventually happen, but VAM surprised us by attaching definite meanings to function words after several foci:

MEANING OF A IS :BLOCK
and
MEANING OF IT IS :B5. [1]

However, it became too expensive to pursue this point, and a pre-defined lexicon was used for future interpretations.

Another interesting question was whether the pre-event focus should be part of the focus of the event, in spite of the hesitation to include this. Experimentally, it seemed to have little effect on what VAM learned. One word which was mis-learned when this part of the focus was expanded (get meant :B2), was corrected (to mean #GOGET) when it was excluded. For the rest of the experiments, we chose to follow the original plan of de-emphasizing the pre-action.

Another question arose with precise definition of the variables u(w), u(c), and u(c,w). The usage of a word could mean the total number of times it is used, or the number of sentences in which it occurs. If we consider the former interpretation, words like the would, indeed, tend to

---

[1] The special characters used at the front of VAM's output distinguish concepts from input vocabulary.

decrease in association values, as they would be the most likely words to re-occur in a single sentence. However, words with direct environmental referents, would seldom be affected. We decided to use the former meaning.

The usage of a concept, u(c), could be the number of times it appears in each focus for all foci, or the number of foci in which it occurs. The greater u(c), the less the correlation, all other terms constant, by virtue of the function itself. Consider the case of an object used in an event in more than one capacity. We would wish to increase its connections to the input words. Therefore, we chose to define u(c) as the number of foci in which a concept occurred.

Thus, we could define u(c,w) as u(c) x u(w) for a given focus-sentence pair. In other words, u(c,w) is increased any time a word is used with a concept, which may be more than once in a given input.

Another experiment involved segmentation and grouping of the input data. Harris's use of pre-segmented idiomatic expressions seemed no less arbitrary than dividing the input strings at lexical boundaries. A child would be unable to recognize word boundaries, whereas he could begin to recognize expressions such as "top-of" as an entity with a meaning of its own. The results in our experiments were

that hyphenated words received the same meaning and associated weights as each of the two used separately, excluding commonly used words like to. With the limited number of instances in the data in which such expressions occurred, it is difficult to say just how representative our results were.

Furthermore, it was found that the data collected from the mother tended to include more than one comment per focus. We tried two means of inputting this data.

The first method was to run the strings together into one large input string. Intuitively, this is a bad idea, considering the evaluation function. An object which is central to the action could appear more than once in several sentences regarding that action, yet its high occurrence would lower its correlation to the concepts in the focus.

The second procedure was to repeat the focus so that each sentence corresponded to the focus equally. In this way, for n input sentences, VAM learns in the same manner as for n occurrences of the identical event, each with a single associated sentence. We ran experimental versions of VAM which did not expand pre-action foci, using data first which was concatenated into one list per focus, as well as being unedited, and then data which was divided into sentences, reusing the focus, with hyphenated idioms. In the case

where foci were not reused, eleven of the learnable words were learned. When reusing foci, VAM learned all of the previously acquired words plus two additional ones.

More verbs were learned in the longer sentences. If two sentences describing the action only mentioned the action in the first, separating the input intc two diminished the correlation.

4.6 Correlation Results

We wished to see what differences would become evident between data drawn from a parent and that of the creator of VAM, who understood the learning processes. For the basic approach, we used pre-segmented, hyphenated versions of the data, with a pre-determined lexicon, reuse of foci, and the above-established interpretation of the correlation function.

Finally, we combined the focus-sentence inputs so that the links were compounded. This provided a larger corpus and more experience for VAM without the necessity of obtaining additional data. Corpus three consisted of 42 words, learned from 120 sentence/focus pairs.

Experimental results for corpus one and corpus two are shown in figures 4.3 and 4.4, respectively. The combined

corpus is in figure 4.5. Multiple entries under "VAM's meaning" indicate that equal weight was attached to two or more concepts.

Other reasons for words being mislearned are attributed to reasons explained by McMaster <1975> and abbreviated in the "reason for error" column:

> 1. The scenes in VAS' experience are such that two concepts c1 and c2 always co-occur. In this case, differences in rounding off weights, or chance ordering of the concepts will decide which of c1 and c2 is chosen. This case is called Uniform Concept Co-occurrence (UCC), and parallels a common error in children.
>
> 2. The corpus contains utterances of the following type: the utterance contains a lexical item whose meaning does not occur in the focus; that is, the utterances contain Misleading Extraneous Words (MEWs).
>
> 3. The correct meaning of a word w is a concept c whose high usage u(c) lowers the value of F(w,c) to a point where c is not chosen as the meaning of w. This case is called Excessive Concept Usage (ECU).
> <McMaster, 1975, pp. 152-153>

Our interpretation of ECU is that it is caused by a low usage of a word in conjunction with a comparatively high usage of the concepts. Therefore, words with very low usage which have a direct environmental referent will probably fall into this category. Adults learning other languages often experience problems with words they have seldom encountered, and children probably do, too.

Results were especially encouraging in that, with

WORD	VAM'S MEANING	CRITERION MEANING	REASON FOR ERROR
BOX	:BLOCK	:BLOCK	
BLOCK	:B1	:BLOCK	ECU
INSIDE	:BOX,:B9, or #INBOX	#INBOX	UCC
TOP-OF	:CUBE	#SUPPORTS	ECU
TRIANGLE	:PYR	:PYR	
GREEN	:GREEN	:GREEN	
RED	:RED	:RED	
GONE	#EXIST	#EXIST or RELOC	
BLUE	:BLUE	:BLUE	
BOOK	:B8	:B8	
BEHIND	:B5,:B3	#FRONT-OF	ECU
FRONT-OF	#TRANSFER	#FRONT-OF	ECU
B7	:B7	:B7	
FALL	#DROP	#DROP	
DROPPED	#LETGO	#LETGO	
TURNED	#TURN	#TURN	
FELL	#DROP	#DROP	
BLACK	:BOX, :B9 OR #INBOX	:BLACK	ECU, UCC
RIGHT	:BOX, :B9 or #INBOX	#BESIDE	ECU
B3	:BOX, :B9 or #INBOX	:B3	ECU
PICKED-UP	#TRANSFER	#TRANSFER or GRASP	
PICK-UP	#TWIST	#TRANSFER or GRASP	ECU
MOVED	#LETGO	AMOVE or PTRANS	ECU
RECTANGLE	#BIGGER, :B1	:BLOCK	ECU
DISAPPEARED	#FRONT-OF	#EXIST	ECU
CUBE	:B3	:CUBE	ECU
SQUARE	#FRONT-OF	:CUBE	ECU

Figure 4.3: Results Using Corpus 1

corpus 3, VAM learned 75% of the learnable words. In general, verbs were seldomed discussed, which would indicate

Word	VAM's meaning	Criterion meaning	Reason for error
BOX	:BOX,:B9, or #INBOX	:BOX	UCC
GREEN	:B3	:GREEN	
B9	:BOX,:B9, or #INBOX	:B9	UCC
INSIDE	:BOX,:B9, or #INBOX	#INBOX	UCC
B3	:B3	:B3	
B1	:B1	:B1	
B6	:B6	:B6	
B7	:B7	:B7	
B4	:B4	:B4	
VAM	#VAM	#VAM	
MOVED	RELOC	RELOC	
DISAPPEARED	#EXIST	#EXIST	
B8	:B8	:B8	
BLUE	:B8	:BLUE	UCC
B2	:B2	:B2	
B5	:B5	:B5	
DROPPED	#LETGO	#LETGO	
GET	#GOGET	#GOGET	
PUSHED	#TWIST	#TWIST	
KNOCKED	#DROP	#DROP	
THREW	PROPEL	PROPEL	
AROUND	#TURN	#TURN	
LIFTED	#MOVEARM	#MOVEARM	
CUBE	:B1	:CUBE	ECU
TOP-OF	#SUPPORTS	#SUPPORTS	
BLOCK	:BLOCK	:BLOCK	
RED	:B5	:RED	ECU
PYRAMID	:PYR	:PYR	
BLACK	:B7, :B3	:BLACK	ECU
PICKED-UP	:BLUE	#TRANSFER or GRASP	ECU
BEHIND	:B5	#FRONT-OF	ECU
ARM	:B5	:ARM	ECU

Figure 4.4: Results Using Corpus 2

that, if our corpora are representative of what children often hear, this could contribute to why verbs are not

Word	U(w)	VAM's Meaning	Criterion Meaning	Reason for Error
BOX	78	:BLOCK	:BOX	
CUBE	5	:B1	:CUBE	ECU
GREEN	29	:GREEN	:GREEN	
TOP-OF	29	#SUPPORTS	#SUPPORTS	
BLOCK	10	#SUPPORTS	:BLOCK	NEW
RED	25	:RED	:RED	
PYRAMID	7	:B5	:PYR	ECU
B9	2	#GOGET	:B9	ECU
INSIDE	12	#INBOX, :BOX, :B9	#INBOX	
B3	6	:B3	:B3	
B6	6	:B6	:B6	
B1	3	#EXIST	:B1	ECU
B7	6	:B7	:B7	
B4	5	#BIGGER	:B4	ECU
VAM	20	VAM	VAM	
BLACK	14	#INBOX, :BOX, :B9	:BLACK	ECU, UCC
MOVED	9	#MOVEARM	#MOVEARM or RELOC	
DISAPPEARED	4	#EXIST	#EXIST	
GONE	6	#EXIST	#EXIST	
B8	6	#TRANSFER	:B8	ECU
BLUE	21	:BLUE	:BLUE	
BESIDE	2	:B2	#BESIDE	ECU
B2	7	:B2	:B2	
PICKED	8	#TRANSFER	GRASP	UCC
BEHIND	7	:B5	#FRONT-OF	ECU
B5	7	:B5	:B5	
DROPPED	4	#LETGO	#LETGO	
GET	2	#GOGET	#GOGET	

(Continued)

acquired until later.

With corpus 1 VAM did not learn a single binary relationship, but in the other tests, it did learn to associate top-of with #SUPPORTS. #FRONT-OF and #BESIDE were never correctly associated. However, had Winograd's

Word	U(w)	VAM's Meaning	Correct Meaning	Reason for Error
OVER	5	#TWIST	#TWIST	
PUSHED	1	#TWIST	#TWIST	
TURNED	5	#TURN	#TURN	
AROUND	1	#TURN	#TURN	
KNOCKED	1	#DROP	#DROP	
THREW	1	PROPEL	PROPEL	
LIFTED	1	#MOVEARM	#MOVEARM	
ARM	1	:B5	:ARM	ECU
TRIANGLE	20	:PYR	:PYR	
RECTANGULAR	3	#EXIST	:BLOCK	ECU
BOOK	11	:B8	:B8	
PICK	3	#TWIST	#TRANSFER or GRASP	MEW
FALL	1	#DROP	#DROP	
FELL	2	#DROP	#DROP	

Figure 4.5: Results Using Combined Corpus

representation been used rather than Schank's, on the same data, results could have been significantly different. VAM had no trouble learning inside, disappeared or gone, except that #INBOX always co-occurred with :B9 and :BOX, a perfect example of UCC.

In Corpus 1, the word box was used for all objects in the environment except the arm. Correspondingly, VAM associated the word with the concept :BLOCK. In Corpus 2, box and block were used as originally intended. VAM learned the correct correlations in this case. Combining the two, we notice that VAM is somewhat confused.

VAM learns verbs and the permanent attribute of color with less exposure than appears to be required for names of objects. This contradicts what we see in children, but cognitive development may be the major discrepancy between the two. Shapes were not so easily learned.

Finally, in corpus 3, out of 14 verbs and adverbs which could be mapped to verb concepts, 12 were successfully linked to valid concepts.

5
CONCLUSIONS

An attempt has been made here to extend VAS's capabilities to more closely approach that of CLAP in its first strategy. Before continuing to implement strategy I, the VAS/VAM approach should be examined for alternative evaluation procedures. CLAP, itself, should be examined regarding its integrity, noting possible omissions or optional approaches in establishing the desired results.

VAS trivially created a segment list, created a Focal Region and a Focal Structure, and built weighted associations between segments and concepts representing only physical objects, their attributes, and relations between them <McMaster, 1975, pp. 154-155>.

VAM creates a Focal Structure, not from the list of blocks given VAS, but from co-ordinates in three-dimensional space and from the occurrences of actions in the environment. Thus, VAM is capable of building weighted

associations to actions. Due to pre-segmentation, as well as the absence of time in VAM's environment, there is no recognition of plural or past-tense forms of the same word or of their being lexically related.

5.1 Extensions to VAM

Ideally, VAM would be connected to a CRT, or even a three-dimensional world, which could be visible to the user. At the present, however, the user must keep track of the placement of objects so as not to try to place a block in mid air, and so on. There is enough intelligence in the programs to make such a block fall, but it must have a level surface directly under it on which to rest.

Both VAS and VAM are programmed in the MTS version of Maclisp, which has no interface to other languages or CRT hardware in the present computer environment. Movement in the environment can only be performed by the programmer, with a pre-determined series of events. Even were VAM's environment a dynamic one, unless the robot were capable of initiating movement on the environment, the system would be lacking a very important source of information which a child has. Perhaps this importance, however, is diminished when all cognitive processes are pre-specified.

The evaluation of the weight on a word-concept link deserves some critical consideration. It is uncertain how any house-keeping will occur during the learning of segmentation, at the same time maintaining function words in the lexicon. By re-evaluating each word of an utterance when it appears in conjunction with a focus, without considering the meaning the word already has and the word linked most closely with each concept, CLAP could be guilty of using less information than that available to the child. With the static function used, the order in which words were acquired had no effect on the correlation. Perhaps it should.

Likewise, one needs to consider the value of salience to CLAP, and devise some criterion for an orienting response. Salience could be built in to the function itself, or the means of constructing the focal list. It appears that actions should draw CLAP's attention more than a static scene. There are many more factors to be sorted out in conjunction with future research in the area of perception. Finally, the fuction of the parameter m in the formula, will have to be established before its value can be

accurately determined.[1]

There is another action which comes to mind as an immediate expansion of VAM, if a sufficiently animated cartoon could portray it to the person providing verbal input. HIT is PROPELing an object as a result of MOVEing the arm, while the arm does not contain it:

```
VAM, PROPEL, k
k, PTRANS, k, a(k), r(k), o(k), a(o(k))
VAM, MOVE, arm.
```

This is a nice event to consider if one has an animated environment, but it is almost impossible to represent the verb "hit" in a series of static scenes, in a manner which prompts relevant input data.

Unlike a child, VAM has a limited external environment. There is no input corresponding to the stimuli of needs, wants, pain, emotions, perception (other than visual), remembering, cognitive development, and complex reasoning. Some of these things are programmed into the system. For instance, VAM is programmed to accept each input as data, and to act upon it, whereas a child may be distracted or ignore what she sees or hears.

---

[1] An alternative correlation function, which takes into consideration the focus and sentence totals (and thus the number of foci/sentences in which the concept/word did not occur), was tested for a few cases. Results are shown in Appendix B.

It is difficult to say just how much confidence one can place in CLAP based on the efforts of VAS and VAM. The capabilities of CLAP are much greater, even during Strategy 1.

5.2 Towards the Implemetation of Strategy 1

Neither VAM nor VAS did any house-keeping on words which were not used. This process would be closely linked to segmentation procedures, whether those outlined by McMaster are to be implemented, or another is adapted. As segments are replaced by those with greater meanings, the nonsense segments must disappear from the vocabulary. To complete Strategy 1, work will need to be done on the segmentation, because both VAS and VAM receive pre-segmented input. This pre-segmentation can be disadvantageous in the association of morphemes with their concepts. If segmentation automatically occurs at word boundaries, one would never learn to recognize "on-top-of" as a single concept. Likewise, words often contain several morphemes, such as tense markers, plurality, etc.

Disapproval as an input has not yet been implemented. Neither is mental maturing a part of VAS or VAM. Similarly, no attempt has been made to allow an action to stimulate

CLAP to output. Both of these are specified in CLAP's first strategy.

It will be interesting to see just how much of a tool a fully-implemented Strategy 1 will be for future strategies, and to what extent each strategy can and must build upon the preceding one. Hopefully, as CLAP becomes a reality, one will be able to observe many child-like problems and successes.

There is no claim made here as to the completeness of CLAP as a model, in recognition of the fact that the later strategies are skeletal at present. However, in CLAP, we finally have an outline of all the necessary processes, based on current research, of a program which can acquire a language.

References

REFERENCES

Anderson, John R., and Gordon H. Bower, Human Associative Memory. Toronto: John Wiley and Sons, 1973.

Block, H.D., J.Moulton, and G.M. Robinson, "Natural Language Acquisition by a Robot", International Journal of Man-Machine Studies. 7, 1975, p. 571-608.

Bowerman, Melissa, Early Syntactic Development. Cambridge, England: Cambridge University Press, 1973.

Brown, Roger W., A First Language: The Early Stages. Cambridge, Massachusetts: Harvard University Press, 1973.

Carroll, John Bissell, Language and Thought. Englewood Cliffs, New Jersey: Prentice-Hall, 1964.

Chomsky, Noam, Aspects of the Theory of Syntax. Cambridge, Massachusetts: MIT Press, 1965.

Dale, Philip S., Language Development: Structure and Function. Hinsdale, Illinois: The Dryden Press, Inc., 1972.

Deese, James Earle, Psycholinguistics. Boston: Allyn and Bacon, 1970.

Derwing, Bruce L., Transformational Grammar and Language Acquisition. Bloomington, Indiana: Indiana University. Thesis, Xerox copy: Ann Arbor, University Microfilms, 1970.

Edwards, Derek, "Sensory-Motor Intelligence and Semantic Relations in Early Child Grammar", Cognition. 2, 1974, pp. 395-434.

Flavell, J.H., The Developmental Psychology of Jean Piaget. Toronto: D. Van Nostrand Company, Inc., 1963.

Harris, Larry R., A Model for Adaptive Problem Solving Applied to Natural Language Acquisition. Ithaca, New York: Cornell University, 1972.

Hunt, Earl, "What Kind of Computer is Man?", Cognitive Psychology. 2, 1971, pp. 57-98.

Jordan, Sarah R., Learning to Use Contextual Patterns in Language Processing. Technical Report No. 152, Computer Sciences Departmen University of Wisconsin, 1972.

Kelley, K.L., Early Syntactic Acquisition. Rand Corporation, Report No. P-3719, 1967.

King, Janet, Ian McMaster, and Jeffrey R. Sampson, "Computer Acquisition of Natural Language: Experimental Tests of a Proposed System", Proceedings of the International Conference on Computational Linguistics, Ottawa, 1976.

McMaster, Ian, Acquisition of Natural Language by a Computer. MSc Thesis, Department of Computing Science, University of Alberta, Edmonton, Alberta, 1975.

McMaster, I, J.R. Sampson, and J.E. King, "Computer Acquisition of Natural Language: A Review and Prospectus", International Journal of Man-Machine Studies. 8, 1976, in press.

Macnamara, John, "Cognitive Basis of Language Learning in Infants", Psychological Review. 79, 1972, pp. 1-13.

McNeill, David, "Developmental Psycholinguistics", in Frank Smith and George A. Miller (Eds.), The Genesis of Language. Cambridge, Massachusetts: MIT Press, 1966, pp. 15-84.

Miller, P.L., "A Locally-Organized Parser for Spoken Input", Communications of the ACM. 17, 1974, pp. 621-630.

Nelson, Katherine, "Concept, Word, and Sentence: Interrelations in Acquisition and Development", Psychological Review. 81, 1974, pp. 267-285.

Pearson, E.S., and H.O.Hartley (Eds.), Biometrika Tables for Statisticians, Volume I. Cambridge: University Press, 1956.

Raphael, Bertram, "Artificial Intelligence", Computer. May, 1973, p. 9.

Reeker, L., A Problem-Solving Theory of Syntactic Acquisition. Ph.D. Thesis in Computing Science, Carnegie-Melon University, 1974.

Richelle, Marc, L'Acquisition du langage. Brussels: Charles Dessart, 1971.

Schank, Roger C., The Development of Conceptual Structures in Children. Stanford AI Project, Memo AIM-203, Stanford University, Stanford, California, 1973a.

Schank, Roger C., The Fourteen Primitive Actions and Their Inferences. Stanford AI Project, Memo AIM-183, Stanford University, Stanford, California, 1973b.

Schank, Roger C., "Identification of Conceptualizations Underlying Natural Language", in Roger C. Schank and Kenneth M. Colby (Eds.), Computer Models of Thought and Language. San Francisco: Freeman and Company, 1973c., pp. 187-247.

Schank, Roger C., and Lawrence G. Tesler, "Margie: Memory, Analysis, Response Generation, and Inference on English", Proceedings of the Third IJCAI, 1973d, pp. 255-269.

Schwarcz R.M., "Steps Towards a Model of Linguistic Performance: A Preliminary Sketch", Mechanical Translation. 10, 1967, pp. 39-52.

Slobin, Dan I., "The Acquisition of Russian as a Native Language", in Frank Smith and George A. Miller (Eds.), The Genesis of Language. Cambridge, Massachusetts: MIT Press, 1966, pp. 129-148.

Winograd, Terry, Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. MIT AI Laboratory, Cambridge, Massachusetts, 1971.

Winograd, Terry, Understanding Natural Language. New York: Academic Press, 1972.

APPENDIX.A:
THE CORPORA FOR VAM

Corpus 1

(SEE LOOK AT THAT)
(HMM WHAT'S THAT)
(IS THAT A TRIANGLE INSIDE A BOX, EH)

(YES NOW WE HAVE THIS TRIANGLE THAT'S DISAPPEARED FROM THE TOP OF THIS BOX, THE SQUARE BOX)
(AND THE TALL BLUE BOX IS BEHIND IT)
(THERE'S THIS BLACK BOX TO THE RIGHT OF THAT)

(LOOK, THERE'S NOW THIS BOX ON TOP OF THAT BOX, SEE)
(AND ALL THE OTHER BOXES ARE IN THE SAME PLACE EXCEPT FOR THIS LITTLE ONE BEHIND HERE)
(HMM, SEE THAT RED ONE)
(NO ANITA DOESN'T SEE IT)

(NOW THE BLUE BOX IS GONE FROM INSIDE OF THE BOX TO THE TOP OF THE RED BOX)
(LOOKS LIKE A THREE-SIDED FIGURE)

(AND THIS BOX IS HOLLOW WITH A LITTLE BOX INSIDE THE BLACK BOX)

(WELL THIS BOX HAS BEEN TURNED OVER ON ITS SIDE THAT'S WHAT'S HAPPENED)
(I THINK ALL THE OTHER ONES ARE IN THE SAME PLACE)

(AND THESE ARE ALL DIFFERENT SIZES)
(ONE BOX IS GONE FROM THE PICTURE)
(THAT'S THE AH TALLEST RECTANGULAR FIGURE THE ONE THAT'S GONE)
(WELL IT'S POINTED, WHAT IS IT)
(TRIANGLE, SORRY I HAD IT WRONG)

(OH THERE, LOOK, THERE'S SOME LITTLE BOXES THERE)
(YEAH AND THERE'S A GREEN ONE ON TOP OF A RED ONE)
(AND ON TOP OF THAT ONE IS A BLUE TRIANGULAR BOX)
(GREEN BOX AND A BLUE THIN BOX BEHIND THAT)
(AND A BLACK BOX WITH A GREEN BOX INSIDE)

(NOW LOOK, ANITA, THAT'S HANGING IN THE AIR, THIS ONE)
(IT'S NOT SITTING ON ANYTHING)

(IT'S BEING MOVED)
(YEAH, THAT LITTLE TINY GREEN ONE)
(THAT LITTLE TINY GREEN ONE, YEA)
(DID YOU KNOW THAT THIS WAS A SQUARE BOX ANITA)
(THIS ONE HERE, THIS GREEN ONE)
(I THINK YOU'D CALL THAT A CUBE)
(ANITA DID YOU KNOW THAT THEY USED TO BUILD PYRAMIDS LIKE IN TRIANGLES)

(NOW THAT LITTLE GREEN GUY HAS BEEN DROPPED FROM THE MIDDLE OF THE AIR AND SITTING DOWN)
(IF ANYTHING ELSE HAPPENED I DON'T KNOW)

(THE RED ONE THE RED TRIANGLE YOUR BOX THAT HAS DISAPPEARED HAS NOW REAPPEARED IN THE BLACK BOX)
(AND ALL THE OTHER BOXES ARE IN THE SAME PLACE)

(NOW THE BLUE BOX IS TURNED SIDEWAYS SO THE THINNER PART IS FACING TOWARDS THE GREEN BOX WHICH IS IN FRONT OF IT)
(LOOKS LIKE A BOOK)
(AND NOW WE'VE PUT THE SMALL RED BOX ON TOP OF THE GREEN BOX)
(NOW THE LITTLE GREEN TRIANGLE IS ON TOP OF THE RED BOX WHICH IS ON TOP OF THE GREEN BOX WHICH IS IN FRONT OF THE RECTANGLE THAT LOOKS LIKE A BOOK)
(AND ALL THE OTHER FIGURES LOOK THE SAME)

(NOW WE'VE GOT THE TRIANGLE IN MIDAIR)
(THE COMPUTER KNOWS IT'S HOLDING IT UP THERE IN MIDAIR EH)
(NOW THERE'S NOTHING MORE INTERESTING ABOUT THAT)
(WE'LL SEE WHERE THAT THING GOES)
(IT'S GONE)
(LOOK WHAT HAPPENED IT FELL DOWN)
(SEE, IT FELL DOWN)
(HEY THAT'S A RED BOX)
(OH NO LOOK AT THAT)
(AND ALL THE OTHER BOXES ARE THE SAME)
(THE RED BOX IS SHORT NOW BECAUSE IT'S NOT STANDING UP ON END ANYMORE)

(AND NOW WE PUT THE GREEN BOX ON TOP OF THE RED BOX AND MOVE THE OOP THE GREEN OTHER GREEN BOX FROM THE OUTSIDE AND PUT THAT IN IT IN FRONT OF THE BLUE BOOK)
(AND THE TRIANGLE'S STILL IN THE BLACK BOX)

(NOW WE'VE PICKED UP THE BOOK THE, COMPUTER'S PICKED UP THE BOOK THAT IS)

(AND MOVED IT IN FRONT OF THE BLACK BOX)
(LOOK NOW WE'VE JUST LET'S SEE)
(ANITA ANITA WHAT'S THAT)
(THAT'S A BOX)

(THAT'S A BOX)

(NOW THE BOOK IS ON ITS BACK)
(SEE THE BLUE BOOK FALL)

(NOW THE COMPUTER'S GONNA PICK UP THE BOOK)
(YEAH IT'S GONNA PICK UP THE BOOK)

(IT DIDN'T PICK UP THE BOOK)
(IT DID SOMETHING ELSE)
(NOW WE'VE GOT THE TRIANGLE ON TOP OF THE BOX, THEREFORE)
(THEY'RE BOTH BLUE)
(IT'S GOT A RED TRIANGLE IN THE BLACK BOX)

(NOW THE RED THIN TALL TRIANGLE IS GONNA BE PICKED UP)
(NO THERE'S NOTHING NEW IN THERE)

(NOW WE'RE PUTTING IT SOMEWHERE)

(WE PUT THE RED TALL TRIANGLE BEHIND THE BLUE FAT TRIANGLE)
(IT'S THE FATTEST ONE)

(NOW THE SMALLEST TINY TRIANGLE THE GREEN ONE IS BESIDE THE BLUE FAT TRIANGLE WHICH IS IN FRONT OF THE SKINNY TRIANGLE)

(OK NOW THE SMALL RED BLOCK IS GONE FROM ON TOP OF THE GREEN BOX)
(THAT'S THE TINIEST BLOCK)

(NOW HE'S TAKING THE GREEN BOX B3 AND PUTTING IT INSIDE THE BLACK BOX)
(IT'S ABOUT THE SAME SIZE AS B7 WHICH IS SITTING ON TOP OF THE RED LONG RECTANGULAR BOX)

(NOW WE'RE TAKING THE FAT BLUE TRIANGLE AND PUTTING IT ON TOP OF THE GREEN BOX WHICH IS ON TOP-OF THE RED LONG RECTANGULAR BOX)

(NOW WHAT HAPPENED)
(THIS ONE WAS GONE AND NOW IT'S ON TOP I THINK)
(YEA I'M RIGHT)
(THE LITTLE TINY RED BLOCK ON TOP-OF THE GREEN BOX INSIDE THE BLACK BOX BEHIND THE THIN BOX)

(NOW WE'RE TAKING THE TALL TRIANGLE AND PUT IT ON TOP-OF THE SMALL BLOCK INSIDE THE BLACK BOX AND LEFT EVERYTHING ELSE THE SAME)

Corpus 2

(THE COMPUTER'S ARM IS ABOVE B5 , THE RED PYRAMID)
(WE HAVE NINE BLOCKS THREE ARE GREEN TWO ARE BLUE AND THREE ARE RED AND A BLACK BOX)

(VAM MOVED B5 BEHIND B3)

(VAM MOVED B7 TO THE TOP OF THE OTHER GREEN BLOCK , B3)

(VAM MOVED B4 TO THE TOP OF B6)

(B7 IS INSIDE THE BLACK BOX)

(VAM TURNED B7 OVER)

(B5 , THE RED PYRAMID , DISAPPEARED)

(VAM WENT TO GET B2)

(VAM LIFTED B2 , THE LITTLE GREEN PYRAMID , AND MOVED IT)

(VAM DROPPED B2)

(B5 IS BACK AGAIN , INSIDE THE BLACK BOX)

(B8 GOT TURNED AROUND BY VAM)

(THE RED CUBE IS ON TOP OF THE GREEN CUBE , B3)

(VAM PUT B2 ON TOP OF B1)

(VAM PICKED UP B4)

(VAM THREW B4 AWAY)

(VAM KNOCKED B6 OVER)

(B7 IS MOVED ON TOP OF B6)

(VAM PICKED UP B8 , THE BLUE BLOCK)

(B8 GOT DROPPED)

(B8 TURNED OVER ON ITS SIDE)

(VAM PUSHED THE BLUE BLOCK OVER SOME MORE)

(B4 , THE BLUE PYRAMID , GOT PUT ON TOP OF THE BLUE BLOCK , B8)

(VAM WENT TO GET B5 WHICH IS STILL INSIDE B9)

(VAM PICKED UP B5 AND MOVED IT OUT OF THE BOX)

(VAM DROPPED B5 BEHIND THE BLUE PYRAMID)

(VAM PICKED UP B2 AND SET B2 DOWN BESIDE THE BLUE PYRAMID ON TOP OF THE BLUE BLOCK B8)

(B1 DISAPPEARED)

(VAM MOVED B3 , ONE OF THE GREEN CUBES , INSIDE THE BLACK BOX)

(VAM STACKED B4 ON TOP OF B7 WHICH IS ON TOP OF B6)

(B1 IS BACK ON TOP OF B3 WHICH IS STILL INSIDE B9)

(NOW THE RED PYRAMID IS ON TOP OF THE TINY RED BLOCK WHICH IS SITTING ON TOP OF THE GREEN CUBE IN THE BOX)

APPENDIX B: PRELIMINARY RESULTS WITH AN ALTERNATIVE CORRELATION FUNCTION

An alternative correlation function was tested on five words from the combined lexicon. Preliminary results indicate that this method may be superior to the function, F, used for both VAS and VAM, and described in section 3.2.4.

The formula assumes u(c), u(w), and u(c,w), are Bernouilli random variables, where:

u(c) = 1 if the concept, c, appears in a given focus
0 otherwise

u(w) = 1 if the word, w, appears in a given sentence
0 otherwise

u(c,w) = 1 if both c and w occurred in an input pair
0 otherwise

With n being the number of input sentence/focus pairs, the correlation was computed as being:

$$[n\,u(c,w) - u(c)\,u(w)] \;/\; u(c)\,u(w)\,[n - u(c)]\,[n - u(w)]$$

VAM learned three of the five words with VAS's formula. With the new correlation, it would have learned four.

The results are shown in the following figures, which list those concepts ranked among the top five by at least one of the functions. From these results, one can see that there are some differences in the ordering of the top-ranked associations of the two formulae.

CONCEPT	OLD CORRELATION	OLD RANK	NEW CORRELATION	NEW RANK
#SUPPORTS	6.96000	1	.17056	1
:B3	6.64000	2	.15909	3
:B1	6.06800	3	.15959	2
:CUBE	6.58200	4	.15196	5
#TRANSFER	6.53500	5	.15193	6
:GREEN	6.20901	6	.15685	4

Figure B.1: Comparisons of the Word Block

CONCEPT	OLD CORRELATION	OLD RANK	NEW CORRELATION	NEW RANK
#TRANSFER	2.535	1	.11560	6
#TURN	1.92999	2	.26881	2
:B8	1.92000	3	.28098	1
#WATCH	1.900001	4	.04415	12
#DROP	1.719999	5	.09197	8
#LETGO	1.719999	5	.09197	8
:BLUE	0.240005	17	.24526	3
:BLUE	0.240005	18	.24526	3
:BLOCK	-5.849996	22	.14721	5

Figure B.2: Comparisons of the Word B8

CONCEPT	OLD CORRELATION	OLD RANK	NEW CORRELATION	NEW RANK
#MOVEARM	5.439997	1	.30441	1
AMOVE	5.420001	2	.16042	6
:LINE	5.280006	3	.15500	7
:ARM	5.280006	3	.15500	7
#VAM	5.016670	4	.18256	5
:GREEN	4.580005	5	.12902	8
#FRONT-OF	3.813331	9	.29963	2
PTRANS	1.410009	24	.20513	4

Figure B.3: Comparisons of the Word Moved

CONCEPT	OLD CORRELATION	OLD RANK	NEW CORRELATION	NEW RANK
#VAM	25.60499	1	.39642	1
AMOVE	24.10398	2	.37411	2
:ARM	23.93598	3	.36515	3
:LINE	23.93598	3	.36515	3
:BLACK	21.68348	4	.24878	5
PTRANS	21.06899	5	.22788	6
#TRANSFER	19.84199	6	.32551	4

Figure B.4: Comparisons of the Word VAM

CONCEPT	OLD CORRELATION	OLD RANK	NEW CORRELATION	NEW RANK
:PYR	20.74400	1	.31001	1
RELOC	17.26400	2	.16482	4
#SUPPORTS	15.79999	3	.15167	6
PTRANS	15.21649	4	.08150	13
:RED	15.18549	5	.19352	3
:B5	13.22799	9	.20567	2
#BIGGER	8.7400	15	.16087	5

Figure B.5: Comparisons of the Word Triangle

One advantage of the second correlation function is that one can find its confidence intervals. Using the confidence coefficient of 95%, the results of the probable meaning for each word was as follows:

TRIANGLE-:PYR	+.16 ≤ .31001 ≤ +.46
VAM-#VAM	+.26 ≤ .39642 ≤ +.54
MOVED-#MOVEARM	+.16 ≤ .30441 ≤ +.46
B8-:B8	+.15 ≤ .28098 ≤ +.45
BLOCK-#SUPPORTS	-.03 ≤ .17056 ≤ +.30

(These ranges are approximations only, derived from a table which assumes bivariate normal distributed random variables <Pearson and Hartley, 1956, p. 140>). This indicates that VAM could not place great confidence in these meanings, especially the meaning for block. More work must be done in

this area before any concrete conclusions can be drawn.

APPENDIX C: ABBREVIATIONS USED

Abbreviation		Page Introduced
ACT		17-18
A.I.	Artificial Intelligence	1
ATN	Augmented Transition Network	43
CLAP	Comprehensive Language Acquisition Program	3
CRT	Cathode Ray Tube terminal	65
ECU	Excessive Concept Usage	63,88
LAD	Language Acquisition Device	11
MEW	Misleading Extraneous Words	88
STM	Short Term Memory	45
UCC	Uniform Concept Co-Occurrence	88
VAM	Verbal Acquisition Module	4
VAS	Vocabulary Acquisition System	4